Review of mathematical statistics

Population

Sampling process ↓ ↑ Statistical Inference

Sample

- Population
    - Model specification
        - Parametric and non-parametric models
        - How to define a model?

- Sampling process
    - **Random sample**: Independent and identically distributed observations
    - Other sampling processes: stratified sampling, cluster sampling or non-random processes like convenience sampling or snowball (you ask the participants to provide you with names of those that will be able to provide you with important information) ….
    - Understanding variability
- Statistical inference: The role of uncertainty
- Parametric and non-parametric inference: Population $X \sim f(x|\theta)$
- If the density (probability) function $f(.)$ is known (and θ is unknown) we face a parametric inference problem. If $f(.)$ (and possibly θ) is unknown we face a non-parametric problem.

- **Parametric inference**: Population $X \sim f(x|\theta)$
  - Parameter space Θ - Depends on the chosen model (and possibly on additional information)
  - Support set of a distribution $\rightarrow D_x = \{x : f(x|\theta) > 0, \theta \in \Theta\}$
    - Sample space of X and support set of the distribution. The use of an indicator function
    - Example: $X \sim Po(\theta)$

$$f_X(x) = \begin{cases} \dfrac{e^{-\theta}\,\theta^x}{x!} & x = 0,1,2,\cdots \\ 0 & \text{otherwise} \end{cases} \quad \text{or} \quad f_X(x) = \frac{e^{-\theta}\,\theta^x}{x!} I_{\{0,1,2,\cdots\}}(x)$$

- **Random sample** $(X_1, X_2, \cdots, X_n)$
  - Sample space
  - Sample distribution (this is a central concept in statistics)
  - Examples

- **Statistic**
  - Definition: Real valued or vector-valued function of the random sample. The domain of the function is the sample space
  - Sampling distribution of statistic
  - Observed value of a statistic
  - Examples
- How to get the sampling distribution of a statistic?
  - General approach: $F_{\mathbf{X}}(t) = \Pr\left(T\left(X_1, X_2, \cdots, X_n\right) \le t\right)$
  - Theoretical results – most of them proved using the moment generating function of X (the characteristic function)
  - Approximate procedures
    - Central limit theorem
    - Monte-Carlo simulation (to be developed latter)

- Examples
  - Sample average from a normal population with known mean and variance;
  - Sampling distribution of $T = \sum_{i=1}^{n} X_i$ when we are sampling from a Bernoulli population

- **Sample moments**
  - k-th sample moment about 0: $M'_k = (1/n)\sum_{i=1}^{n} X_i^k$
    - Sample mean $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$
  - k-th sample moment about $\bar{X}$: $M_k = (1/n)\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$
    - Sample variance

$$M_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

- Sample moments *versus* population moments

- Some results (we assume that population moments exist)
  - Sample mean

$$E(\bar{X}) = E(X) = \mu \ ; \qquad \operatorname{var}(\bar{X}) = \frac{\operatorname{var}(X)}{n} = \frac{\sigma^2}{n}.$$

  - Central limit theorem

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \overset{\circ}{\sim} n(0,1)$$

  - Sample variance

$$E(M_2) = \frac{n-1}{n}\sigma^2 \qquad \operatorname{var}(M_2) = \frac{\mu_4-\mu_2^2}{n} - 2\frac{\mu_4-2\mu_2^2}{n^2} + \frac{\mu_4-3\mu_2^2}{n^3} \text{ with } \mu_k = E(X-\mu)^k$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \qquad n>1$$

$$E(S^2) = \sigma^2 \qquad \operatorname{var}(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$$

Order statistics

- **Definition:** The **order statistics** of a random sample $(X_1, X_2, \cdots, X_n)$ are the sample values placed in ascending order. They are denoted by $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ or by $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ or $Y_1 \le Y_2 \le \cdots \le Y_n$
- **Comments:**
  - Unlike the random sample itself, the order statistics are **not independent**. If $Y_j > y$ then $Y_s > y$ for $s > k$.
  - The sample minimum and the sample maximum are examples of order statistics.
  - Remember that the sample median is defined to be the middle order statistic if n is odd ($Y_{(n+1)/2}$) or the average of the middle two order statistics if n is even ($0.5 \times Y_{n/2} + 0.5 \times Y_{1+n/2}$).

- Marginal cumulative distribution of the *r*-th order statistic: Let $(X_1, X_2, \cdots, X_n)$ denote a random sample of size n from a population with cumulative distribution function $F_X(x)$. The marginal cumulative distribution will be $F_{Y_r}(y) = \sum_{j=r}^{n} \binom{n}{j} (F_X(y))^j (1 - F_X(y))^{n-j}$

  Proof:

$$
\begin{aligned}
F_{Y_r}(y) &= \Pr(Y_r \le y) \\
&= \Pr(Y_r \le y \wedge Y_{r+1} > y) + \Pr(Y_{r+1} \le y \wedge Y_{r+2} > y) + \cdots + \Pr(Y_{n-1} \le y \wedge Y_n > y) + \Pr(Y_n \le y) \\
&= \binom{n}{r} (F_X(y))^r (1 - F_X(y))^{n-r} + \binom{n}{r+1} (F_X(y))^{r+1} (1 - F_X(y))^{n-(r+1)} + \\
&\quad + \cdots + \binom{n}{n-1} (F_X(y))^{n-1} (1 - F_X(y))^{n-(n-1)} + \binom{n}{n} (F_X(y))^n (1 - F_X(y))^{n-n} \\
&= \sum_{j=r}^{n} \binom{n}{j} (F_X(y))^j (1 - F_X(y))^{n-j}
\end{aligned}
$$

- If X is a continuous random variable the density function of the *r*-th order statistic will be

$$f_{Y_r}(y) = \frac{n!}{(r-1)!1!(n-r)!}\left(F_X(y)\right)^{r-1}\left(1-F_X(y)\right)^{n-r} f_X(y)$$

Proof: see Casella and Berger

- Example: Let us consider a continuous random variable following an exponential distribution with mean θ and a sample of size 5. The density function of the sample median will be

$$f_{Y_3}(y) = \frac{5!}{2!1!2!}\left(1-e^{-y/\theta}\right)^2\left(e^{-y/\theta}\right)^2\theta^{-1}e^{-y/\theta} = 30\,\theta^{-1}\left(1-e^{-y/\theta}\right)^2 e^{-3y/\theta} \qquad y>0$$

- Let X be a continuous random variable with distribution function $F(x)$ and density $f(x)$. $F(x)$ is strictly monotone for $0<F(x)<1$, and let m be the population median (m is the unique solution of $F(m)=1/2$). Let M be the sample median. Then, it can be proved that M is asymptotically distributed as a normal variable with mean m and variance $(4n\,f(m)^2)^{-1}$, i.e.

$$(M-m)\left(2\,f(m)\sqrt{n}\right)\overset{\circ}{\sim} n(0;1)$$

POINT ESTIMATION

- We are in the core of parametric inference i.e. we have a model and we want to estimate the unknown parameter(s), i.e. $X \sim f(x|\theta), \ \theta \in \Theta$ where $f(.)$ is a known density (probability) function and θ is an unknown parameter.
- In real world we could also consider that our knowledge of $f(.)$ is questionable (i.e. that we are considering a Pareto distribution when the real one is, for instance, lognormal), but, at this stage, we will not proceed in such direction.
- Regardless of how the parameter is estimated it is extremely unlikely that the estimated model will match exactly the true model.
- The problem is then to measure the error we will be making. However if we were able to measure the error we could correct our estimate by that amount and have no error at all.
- What we are able to do is to evaluate the procedure that generates the estimate and not the estimate itself. We must distinguish between the **estimator** and the **estimate**.

- Keep in mind that a good procedure can lead to a poor estimate and conversely a poor procedure can originate a good estimate. However good procedures are more likely to produce good estimates than poor procedures.
- This evaluation is done considering the set of results that could have been generated by the procedure and not a particular one.
- **Example**: To illustrate the situation let us consider that we want to estimate the mean θ of a normal population with known variance σ^2 using the mean of a sample of size n.

  The intuitive procedure is to use the sample average, i.e. $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ as estimator or $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ as an estimate.

  The quality of the procedure (the estimator) is evaluated using the sampling distribution of $\bar{X}$. As it is well known, $\bar{X}$ is a random variable that follows a normal distribution with mean θ and variance σ^2/n, i.e., $\dfrac{\bar{X}-\theta}{\sigma/\sqrt{n}} \sim n(0;1)$

Unbiasedness

- **Definition 12.1**: An estimator $\hat{\theta}$ is **unbiased** if $E(\hat{\theta} \mid \theta) = \theta, \forall \theta \in \Theta$. The **bias** $bias_{\hat{\theta}} = E(\hat{\theta} \mid \theta) - \theta$.

- Comments:

  - The point is to verify the equality $\forall \theta \in \Theta$ (see example 2)
  - The bias depends on the estimator being used but also on the particular value of θ.
  - An estimator with a positive bias tends to overestimate the parameter.

- **Example 1:** Prove that the sample mean is an unbiased estimator for the population mean (assume that the population mean exists).

  Let us denote the population mean by μ.

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

- **Example 2:** Consider a Bernoulli population with mean θ and $T_2 = 0.3$ as an estimator for θ. As it is obvious T_2 is a bad estimator since it does not take into account the sample values. For $\theta = 0.3$, $E(T_2) = \theta$ but T_2 is a biased estimator since the equality $E(T_2) = \theta$ is not true $\forall \theta \in \Theta$.
- **Example 12.4**: A population has an exponential distribution with a mean θ. We want to estimate the population mean using a sample of size 3. Determine the bias of the sample mean and the sample median as estimators of the population mean.

  Sample mean:

  $E(\overline{X}) = \theta$

Sample median

Let T be the sample median. Then

$$f_T(t) = \frac{3!}{1!1!1!}\left(1-e^{-t/\theta}\right)\left(e^{-t/\theta}\right)\theta^{-1}\, e^{-t/\theta} \qquad t>0$$

$$= 6\theta^{-1}\left(1-e^{-t/\theta}\right)e^{-2t/\theta} = 3(2/\theta)e^{-2t/\theta} - 2(3/\theta)e^{-3t/\theta} \qquad t>0$$

$$E(T\mid\theta) = \int_0^\infty t\, f_T(t)\, dt$$

$$= \int_0^\infty t\left(3(2/\theta)e^{-2t/\theta} - 2(3/\theta)e^{-3t/\theta}\right)dt = 3\int_0^\infty t\,(2/\theta)e^{-2t/\theta}\,dt - 2\int_0^\infty t\,(3/\theta)e^{-3t/\theta}\,dt$$

$$= 3\times\frac{\theta}{2} - 2\times\frac{\theta}{3} = \frac{9\theta-4\theta}{6} = \frac{5\theta}{6}$$

The bias is

$bias_{\hat\theta} = -\theta/6$ On average, the estimator underestimates the population mean θ which is not a surprise. Remember that the median of the population is $\theta\ln 2$ ($<\theta$). Note that the sample median is also a biased estimator for the population median.

- **Definition 12.2**: An estimator $\hat{\theta}$ is **asymptotically unbiased** if $\lim_{n\to\infty} E(\hat{\theta}\mid\theta)=\theta, \forall\,\theta\in\Theta$.

- **Example 12.5** $X \sim U(0;\theta)$, sample $(X_1, X_2, \cdots, X_n)$ and $\hat{\theta}=\max X_i$.

$$f_X(x)=1/\theta \qquad 0<x<\theta$$

$$f_{\hat{\theta}}(y)=\frac{n!}{(n-1)!1!0!}(F_X(y))^{n-1}(1-F_X(y))^0 f_X(y)$$

$$=n(y/\theta)^{n-1}(1/\theta)=n\theta^{-n}y^{n-1} \qquad 0<y<\theta$$

$$E(\hat{\theta}\mid\theta)=\int_0^\theta y\, f_{\hat{\theta}}(y)\,dy=\int_0^\theta y\,n\,\theta^{-n}y^{n-1}\,dy=\theta^{-n}\int_0^\theta n\,y^n\,dy$$

$$=\frac{n}{n+1}\theta$$

The estimator is biased but is asymptotically unbiased as

$$\lim_{n\to\infty}E(\hat{\theta}\mid\theta)=\lim_{n\to\infty}\frac{n}{n+1}\theta=\theta$$

- **How to compare 2 unbiased estimators**?
  - Let T and T' be 2 unbiased estimators for the parameter θ. We will say that T is better than T' if $\mathrm{var}(T \mid \theta) \le \mathrm{var}(T' \mid \theta)$, $\forall \theta \in \Theta$ (the inequality has to be strict for, at least, one θ).
  - Example: $X \sim Po(\theta)$ and let us consider $T = \bar{X}$ and $T' = S^2$ as estimators of θ.

$$E(T \mid \theta) = E(\bar{X} \mid \theta) = \theta \qquad E(T' \mid \theta) = E(S^2 \mid \theta) = \mathrm{var}(X \mid \theta) = \theta$$

$$\mathrm{var}(T \mid \theta) = \mathrm{var}(\bar{X} \mid \theta) = \sigma^2 / n = \theta / n$$

$$\mathrm{var}(T' \mid \theta) = \mathrm{var}(S^2 \mid \theta) = \frac{1}{n}\left(\theta + 3\theta^2 - \frac{n-3}{n-1}\theta^2\right) = \frac{\theta}{n} + \frac{1}{n}\left(\frac{3n-3-n+3}{n-1}\right)\theta^2 = \frac{\theta}{n} + \frac{2}{n-1}\theta^2 > \mathrm{var}(T \mid \theta)$$

- **Definition** (CB): An estimator T for $\tau(\theta)$ is **a best unbiased estimator** of $\tau(\theta)$ if it satisfies $E(T \mid \theta) = \tau(\theta)$ for all θ and, for any other estimator W with $E(W \mid \theta) = \tau(\theta)$, we have $\mathrm{var}(T \mid \theta) \le \mathrm{var}(W \mid \theta)$ for all θ. T is also called a uniform minimum variance unbiased estimator (**UMVUE**) of $\tau(\theta)$.

- **Cramér-Rao Inequality applied to unbiased estimators**

Let $(X_1, X_2, \cdots, X_n)$ be a random sample from a population with probability density function $f_X(x \mid \theta)$ and let $T = T(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of $\tau(\theta)$ satisfying

$$\frac{d}{d\theta} E(T \mid \theta) = \int_{D_\mathbf{x}} \frac{\partial}{\partial\theta}\big(T(\mathbf{x}) \times f(\mathbf{x} \mid \theta)\big) d\,\mathbf{x} \text{ and } \operatorname{var}(T \mid \theta) < \infty .$$

Then $\operatorname{var}(T \mid \theta) \geq \dfrac{\left(\dfrac{d}{d\theta}\tau(\theta)\right)^2}{n\,\Im(\theta)}$ where $\Im(\theta) = E\left(\dfrac{\partial}{\partial\theta} \ln f_X(X \mid \theta)\right)^2 = -E\left(\dfrac{\partial^2}{\partial\theta^2} \ln f_X(X \mid \theta)\right)$

- Comments
  - The original Cramér-Rao inequality is proved for any estimator and for non-independent sampling – see Casella and Berger, 2nd edition, page 335.

- $\frac{d}{d\theta}E(T \mid \theta) = \int_{D_{\mathbf{x}}} \frac{\partial}{\partial \theta}\left(T(\mathbf{x}) \times f(\mathbf{x} \mid \theta)\right) d\,\mathbf{x}$. We can swap the derivation (in order to θ) with the integration (in order to x). The set of support of X cannot depend on θ (the uniform density function doesn't fulfill this condition).
- $\operatorname{var}(T \mid \theta) < \infty$: The variance of T should exist.
- When we have an unbiased estimator of θ we can compare its variance with the lower bound given by the Cramér-Rao inequality. If they are equal we have an UMVUE. If not, nothing can be concluded (nothing is said about the possibility that an unbiased estimator with a variance equal to the lower bound exists).
- $\Im(\theta)$ is called Fisher information

- **Example** – Consider a Poisson population with mean θ and show that $\bar{X}$ in an UMVUE estimator for θ.

  We have already shown that $\bar{X}$ is an unbiased estimator for θ and that $\operatorname{var}(\bar{X}) = \theta / n$.

  Let us now calculate the lower bound of the Cramér-Rao inequality.

  $$f_X(x \mid \theta) = \frac{e^{-\theta}\,\theta^x}{x!} \qquad \ln f_X(x \mid \theta) = -\theta + x \ln\theta - \ln(x!)$$

  $$\frac{\partial}{\partial\theta}\ln f_X(x \mid \theta) = -1 + \frac{x}{\theta} \qquad \frac{\partial^2}{\partial\theta^2}\ln f_X(x \mid \theta) = -\frac{x}{\theta^2}$$

  $$\Im(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ln f_X(X \mid \theta)\right) = -E\left(-\frac{X}{\theta^2}\right) = \frac{\theta}{\theta^2} = \frac{1}{\theta} \qquad \tau(\theta) = \theta \qquad \left(\frac{d}{d\theta}\tau(\theta)\right)^2 = 1$$

  The lower bound is then $\dfrac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{n\Im(\theta)} = \dfrac{1}{n/\theta} = \dfrac{\theta}{n}$

  As $\bar{X}$ is an unbiased estimator of θ with a variance equal to the lower bound, we can conclude that $\bar{X}$ in an UMVUE estimator for θ.

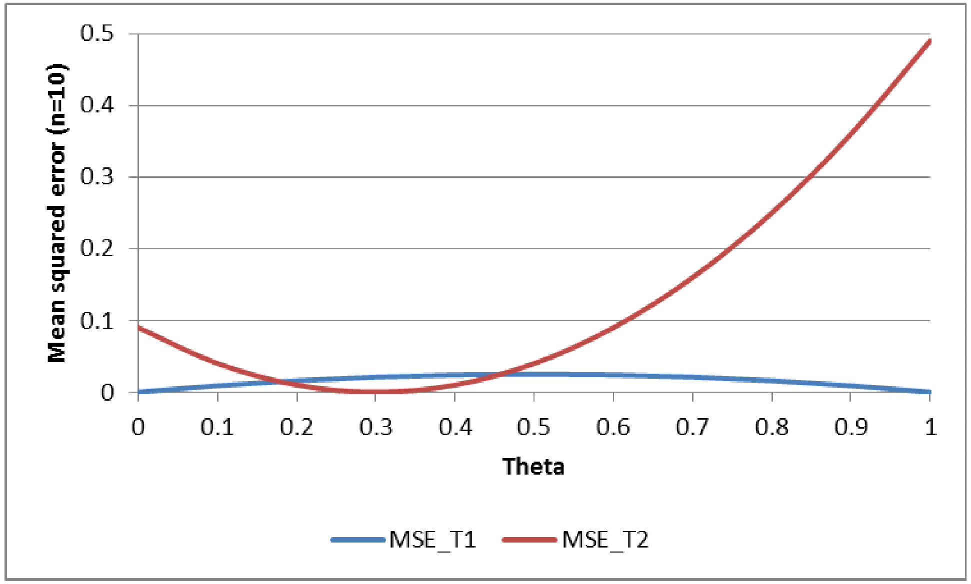
Mean-squared error

- How to compare estimators that are not unbiased?
- **Definition 12.4**: The mean-squared error of an estimator is $MSE_{\hat{\theta}}(\theta) = E\left(\left(\hat{\theta}-\theta\right)^2 \mid \theta\right)$
- The mean-squared error can be rewrite as
  $MSE_{\hat{\theta}}(\theta) = E\left(\left(\hat{\theta}-\theta\right)^2 \mid \theta\right) = \text{var}(\hat{\theta} \mid \theta) + \left(bias_{\hat{\theta}}(\theta)\right)^2$
- Comments
  - The mean-squared error is a function of the true value of the unknown parameter, θ, so that some estimator can perform very well for some values of θ and poorly for other values of θ
  - Using the MSE with an unbiased estimator of θ is the same as using its variance

- **Example**: Let us consider a Bernoulli population with parameter θ and two estimators for θ obtained using a sample of size n: $T_1 = \overline{X}$ and $T_2 = 0.3$. Compare these estimators using their MSE. Comment.

$$MSE_{T_1}(\theta) = E\left((T_1-\theta)^2 \mid \theta\right) = E\left((\overline{X}-\theta)^2 \mid \theta\right) = \text{var}(\overline{X} \mid \theta) + \left(E(\overline{X} \mid \theta)-\theta\right)^2 = \frac{\theta(1-\theta)}{n} + 0 = \frac{\theta(1-\theta)}{n}$$
$$MSE_{T_2}(\theta) = E\left((T_2-\theta)^2 \mid \theta\right) = E\left((0.3-\theta)^2 \mid \theta\right) = (0.3-\theta)^2$$

Although T_2 is an inadequate estimator of θ (the estimator does not take into account the collected sample) we see that $MSE_{T_1}(\theta)$ could be less than $MSE_{T_2}(\theta)$ for some values of θ



- It is convenient to use a qualification criterion before using the MSE and only compare estimator that fulfill such criterion.

Consistency

- **Definition 12.3 –** An estimator is consistent (often called, in this context, weakly consistent) if, for all $\delta > 0$ and any θ, $\lim_{n\to\infty} \Pr(|\hat{\theta}_n - \theta| > \delta) = 0$.
- **Comments:**
  - A sufficient although not necessary condition for weak consistency is that $\lim_{n\to\infty} E(\hat{\theta}_n \mid \theta) = \theta$ and $\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n \mid \theta) = 0$. Such statement can be proved using Markov inequality ($\Pr(|X| \geq a) \leq E(|X|)/a$)[1].
  - Consistency is a property of the sequence of estimators, $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n, \cdots$, and not of the estimator itself.
  - The idea behind consistency is that the estimator has to work well for large samples.

---

[1] $\Pr\left(\left|\hat{\theta}_n - \theta\right| > \delta\right) = \Pr\left(\left(\hat{\theta}_n - \theta\right)^2 > \delta^2\right) \leq \Pr\left(\left(\hat{\theta}_n - \theta\right)^2 \geq \delta^2\right) \leq \frac{E\left(\hat{\theta}_n - \theta\right)^2}{\delta^2} = \frac{\mathrm{var}(\hat{\theta})}{\delta^2} + \frac{\left(E(\hat{\theta}) - \theta\right)^2}{\delta^2}$

- **Example 12.6 –** Prove that, if the variance of a random variable is finite, the sample mean is a consistent estimator of the population mean.

  $E(\overline{X}) = \mu$

  $\text{var}(\overline{X}) = \sigma^2 / n$

  Then

  $$\lim_{n\to\infty} E(\overline{X}) = \lim_{n\to\infty} \mu = \mu$$

  $$\lim_{n\to\infty} \text{var}(\overline{X}) = \lim_{n\to\infty} \sigma^2 / n = 0$$

INTERVAL ESTIMATION

- Unlike **point estimation, interval estimation** leads to a set of values.
- The main point is to associate a level of confidence to such intervals as we will remember.
- **Definition 12.6** – A $100(1-\alpha)\%$ confidence interval for a parameter θ is a pair of random values, *L* and *U*, computed from a random sample such that $\Pr(L \le \theta \le U) \ge 1-\alpha$ for all θ.
- **Comments:**
  - The definition does not uniquely define an interval;
  - When we replace the random variables by their observed values, nothing is said about whether or not the interval encloses θ.
  - The level of confidence is a property of the process and not a property of the particular values obtained
  - Note that the inequality concerns discrete random populations, as we will see.

- How to construct a confidence interval?
  - Not an easy question when considering a general situation
  - Usually we follow the pivotal method
- **Pivotal quantity –** A random variable $Q(X_1, X_2, \cdots, X_n, \theta)$ is a pivotal quantity if the distribution of $Q(X_1, X_2, \cdots, X_n, \theta)$ does not depend on θ.
- **Comments**: The function $Q(X_1, X_2, \cdots, X_n, \theta)$
  - depends only on the sample $(X_1, X_2, \cdots, X_n)$, on θ and, possibly, on some known values;
  - is completely known;
  - usually, is monotonic in θ

- **Pivotal method** (we will assume that $Q(X_1, X_2, \cdots, X_n, \theta)$ follows a continuous distribution)
  - Step 1 – Find q_1 and q_2 such that $\Pr\left(q_1 \le Q(X_1, X_2, \cdots, X_n, \theta) \le q_2\right) = 1 - \alpha$.
  - Step 2 – From $q_1 \le Q(X_1, X_2, \cdots, X_n, \theta) \le q_2$ define L and U such that $q_1 \le Q(X_1, X_2, \cdots, X_n, \theta) \le q_2 \Leftrightarrow L \le \theta \le U$.
  - L and U define a confidence interval for θ. The problem of how to choose the pair q_1 and q_2 remains. Optimally q_1 and q_2 are chosen to minimize the length (or its expected value if such length is random) of the confidence interval. As this task is difficult to fulfill in most situations we can follow a practical approximation and choose q_1 and q_2 such that $\Pr\left(Q(X_1, X_2, \cdots, X_n, \theta) < q_1\right) = \Pr\left(Q(X_1, X_2, \cdots, X_n, \theta) > q_2\right) = \alpha / 2$

- Some well-known pivotal quantities:
  - For normal populations or when we have a large sample some pivotal quantities are well-known for usual situations;
  - For other situations we try to find and estimator $\hat{\theta}$ for θ with a known distribution (independent of θ). If the sample is large enough and the estimator well behaved we can use $\dfrac{\hat{\theta}-E(\hat{\theta})}{\mathrm{var}(\hat{\theta})} \overset{\circ}{\sim} n(0;1)$.

    Note that, as this result is asymptotic, we can use an adequate approximation for $E(\hat{\theta})$ and $\mathrm{var}(\hat{\theta})$

A) Gaussian (normal) populations:

	Pivotal Quantity	Confidence Interval
Mean (known variance)	$Q(X_1, X_2, ..., X_n, \mu) = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$	$\left(\overline{X} - z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} , \overline{X} + z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \right)$
Mean (unknown variance)	$Q(X_1, X_2, ..., X_n, \mu) = \dfrac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}$	$\left(\overline{X} - t_{\alpha/2} \dfrac{S}{\sqrt{n}} , \overline{X} + t_{\alpha/2} \dfrac{S}{\sqrt{n}} \right)$
Variance	$Q(X_1, X_2, ..., X_n, \sigma^2) = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left(\dfrac{(n-1)S^2}{q_2} , \dfrac{(n-1)S^2}{q_1} \right)$
$z_{\alpha/2}: \Phi(z_{\alpha/2}) = 1 - \alpha/2; \quad t_{\alpha/2}: \; P(T_{(n-1)} > t_{\alpha/2}) = \alpha/2; \; q_1, q_2: \; P(Q_{(n-1)} < q_1) = P(Q_{(n-1)} > q_2) = \alpha/2$		

B) Large samples (Confidence interval for the mean):

	Pivotal Quantity	**Confidence Interval (aprox)**
Case 1	$Q(X_1, X_2, ..., X_n, \mu) = \dfrac{\overline{X} - \mu}{\sqrt{\text{var}(\overline{X})}} \overset{\circ}{\sim} N(0,1)$	
Case 2	$Q(X_1, X_2, ..., X_n, \mu) = \dfrac{\overline{X} - \mu}{\sqrt{\widehat{\text{var}}(\overline{X})}} \overset{\circ}{\sim} N(0,1)$	$\left(\overline{X} - z_{\alpha/2} \sqrt{\widehat{\text{var}}(\overline{X})} \, , \, \overline{X} + z_{\alpha/2} \sqrt{\widehat{\text{var}}(\overline{X})} \right)$
Case 3	$Q(X_1, X_2, ..., X_n, \mu) = \dfrac{\overline{X} - \mu}{S / \sqrt{n}} \overset{\circ}{\sim} N(0,1)$	$\left(\overline{X} - z_{\alpha/2} \dfrac{S}{\sqrt{n}} \, , \, \overline{X} + z_{\alpha/2} \dfrac{S}{\sqrt{n}} \right)$
$z_{\alpha/2} : \Phi(z_{\alpha/2}) = 1 - \alpha/2;$		

Bernoulli populations: $X \sim Ber(\theta)$

$$\text{var}(\overline{X}) = \frac{\text{var}(X)}{n} = \frac{\theta(1-\theta)}{n} \quad \text{or} \quad \hat{\text{var}}(\overline{X}) = \frac{\overline{X}(1-\overline{X})}{n}$$

Case 1

$$\left(\frac{\overline{X} - \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}}}{\left(1 + \frac{z_{\alpha/2}^2}{n}\right)} ; \frac{\overline{X} - \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}}}{\left(1 + \frac{z_{\alpha/2}^2}{n}\right)} \right)$$

Case 2

$$\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} ; \overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right)$$

Poisson Populations: $X \sim Po(\theta)$

$$\mathrm{var}(\overline{X}) = \frac{\mathrm{var}(X)}{n} = \frac{\theta}{n} \quad \text{or} \quad \widehat{\mathrm{var}}(\overline{X}) = \frac{\overline{X}}{n}$$

Case 1 -

$$\left(\overline{X} - \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}}{n}} ; \overline{X} - \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}}{n}} \right)$$

Case 2

$$\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}} ; \overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}} \right)$$

TEST OF HYPOTHESES

- Null, H0, and alternative, H1, hypotheses
- The two hypotheses are not treated symmetrically (Neyman-Pearson approach). We do not reject H0 unless there is strong statistical evidence against it.
- The result of a test is the rejection (or not) of the null hypothesis. What so ever the decision is, an error is always possible:
  - Type I error: Rejection of the null when the null is true;
  - Type II error: Not rejecting the null when the null is false.

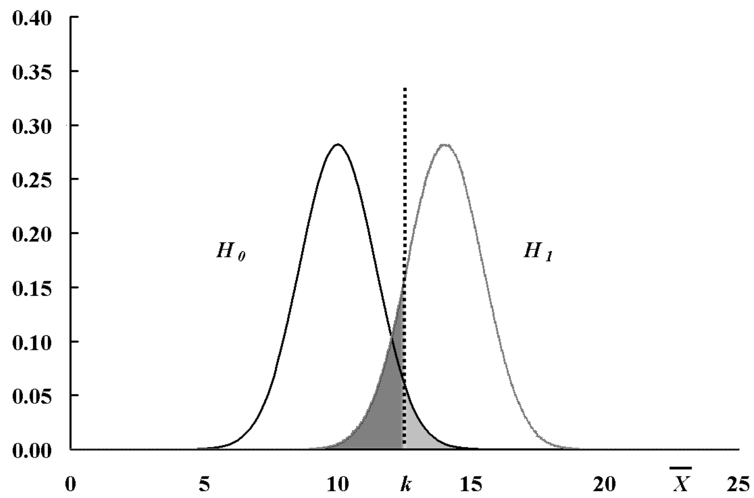
	H_0 **true**	H_0 **false**
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error

- Using a simple example it can be shown that it is not possible to minimize both errors.

  $X \sim N(\mu, \sigma^2)$ with $\sigma^2 = 4$. The test is $H_0 : \mu = 10$ against $H_1 : \mu = 14$.

a) Let us assume that our sample has only one observation and that the rejection region is given by $W = \{x : x > 12.5\}$. Determine the probabilities associated with type 1 and type 2 errors. ($\alpha \approx 0.1056$, $1 - \beta \approx 0.2266$).

b) Show that decreasing the probability of a type 1 error implies increasing the probability of a type 2 error and vice-versa



Note that if we increase the sample size we can reduce simultaneously the probabilities of both errors

- **Definition 12.7** – The significance level of a hypothesis test is the probability of making a Type I error given that the null is true. If it can be in more than one way, the level of significance is the maximum of such probabilities. The significance level is usually denoted by α.
- **Comments**:
  - This definition is conservative since we are considering the worst case;
  - Typically, the worst case is on the boundaries between H0 and H1;
  - Usual values for the level of significance are 1%, **5**% or 10%.
- Using the Neyman-Pearson approach one should control the probability associated with the Type I error, i.e. one must control the significance level of the test, and choose the test with a smaller probability of a Type II error, given the significance level.
- **Comments**:
  - The approach give more importance to the type I error;
  - Such a test is called a most powerful (uniformly most powerful test);
- **Definition 12.8** – A hypotheses test is **uniformly most powerful** (UMP) if no other test exists that has the same or lower significance level and, for a particular value within the alternative hypothesis, has a smaller probability of making a Type II error.

- **Test statistic** –The test statistic is a function of the sample observations with a known distribution under the null. The design of a test procedure looks at all the samples that might have been observed and not at the particular sample that was observed.
- **Rejection region** – The test specification is completed by defining a rejection region. If the observed value of the test statistic falls in the rejection region we will reject the null, otherwise we will not reject the null.
- **How develop a test of hypotheses?**
    - Define the hypotheses H0 and H1 and
    - Choose an adequate significance level
    - Obtain a test statistic and determine the rejection region
    - Calculate the observed value of test statistic and conclude
- Open questions: How to obtain the test statistic and, given the test statistic, how to determine the rejection region?
    - **Theoretical results**: Neyman-Pearson's lemma and Karlin-Rubin theorem
    - **Empirical rule of thumb**: When testing a mean, a variance or a proportion (Bernoulli populations) using the "natural" test statistic the rejection region is on the side of the alternative.

- In most situations a UMP test does not exist, namely when the null hypothesis is an equality and the alternative is both sides ("=" against "$\neq$")

- Some useful results:

Normal populations:

Test about the mean, variance known $$Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim n(0;1)$$

Test about the mean, variance unknown $$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t_{(n-1)} \qquad S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$$

Test about the variance $$Q=\frac{(n-1)S^2}{\sigma^2}\sim \chi^2_{(n-1)} \qquad S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$$

- Large samples:

Test about the mean, variance unknown $$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\overset{\circ}{\sim} n(0;1) \qquad S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$$

Bernoulli population $$Z=\frac{\overline{X}-p_0}{\sqrt{p_0(1-p_0)/n}}\overset{\circ}{\sim} n(0;1)$$

Poisson population $$Z = \frac{\overline{X} - \mu_0}{\sqrt{\mu_0 / n}} \overset{\circ}{\sim} n(0;1)$$

- **Examples 12.13 to 12.15** – Your company has been basing its premium on an assumption that the average claim is 1200. You want to raise the premiums, and a regulator has insisted that you provide evidence that the average now exceeds 1200. To provide such evidence, the following numbers have been obtained:

27 82 115 126 155 161 243 294 340 384 457 680 855 877 974 1193

1340 1884 2558 15743

a) What are the hypotheses for this problem (example 12.13)?

b) Complete the test using the test statistic and rejection region that is promoted in most statistics books ($\alpha = 0.05$). Assume that the population has a normal distribution with standard deviation 3435 (Example 12.14).

c) Determine the probability of making a Type II error when the alternative hypothesis is true with $\mu = 2000$ (Example 12.15).

Answers:

b)

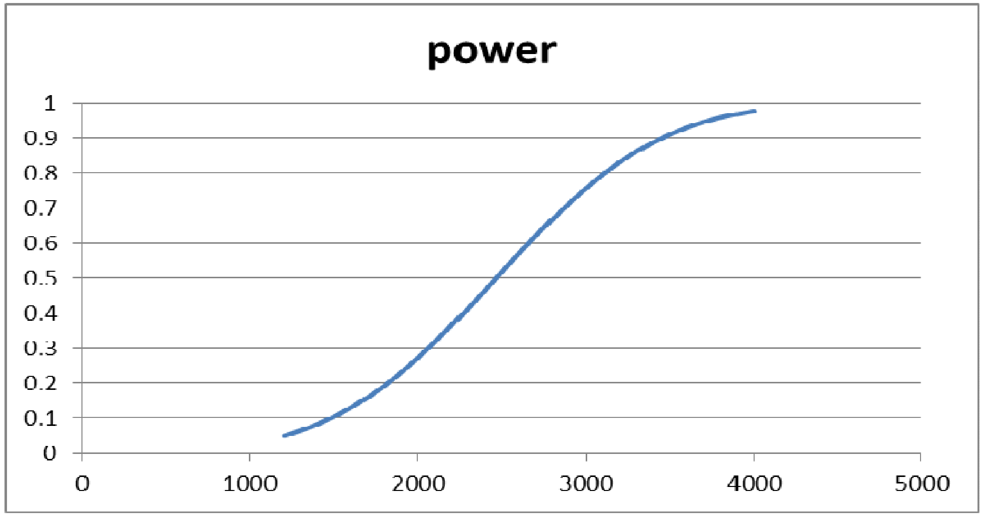
$\bar{x} = 1424.4 \qquad z = (1424.4 - 1200) \times \sqrt{20} / 3435 = 0.292154$

Test (N-P procedure): $\alpha = 0.05 \quad z_\alpha = 1.645$ (one side test)

Rejection region: $W = \{(x_1, x_2, \cdots, x_{20}) : z > 1.645\}$ or $W = \{(x_1, x_2, \cdots, x_{20}) : \bar{x} > 1200 + 1.645 \times 3435 / \sqrt{20}\}$

conclusion: do not reject H0

c) $\Pr(\text{Accept } H_0 \mid \mu = 2000) = \Pr(\bar{X} \le 2463.507 \mid \mu = 2000) = \Pr(Z \le 0.603455) = 0.7269$


power
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
4000
5000

p-values

Under the "classical" (Neyman-Pearson) approach a test will produce a decision on whether or not to reject H_0 for a predetermined value of α.

Sometimes this procedure does not provide the recipient of the result with clear information on the strengthof the evidence against H_0.

A more informative approach is to calculate and quote the P-value of the observed test statistic. This is the significance level of the test statistic, i.e.

- The probability, assuming H_0 is true, of observing a test statistic at least as "extreme" (inconsistent with H_0) as the value observed;

- The significance level that originates a critical value equal to the observed value of the test statistic.
- If α is greater than the p-value we reject H_0 and if α is smaller than the significance level we do not reject H_0

- **Definition 12.9** – For a hypothesis test, the p-value is the probability that the test statistic takes on a value that is less in agreement with the null hypothesis than the value obtained from the sample. Tests conducted at a significance level that is greater than the p-value will lead to a rejection of the null hypothesis, while tests conducted at a significance level that is smaller than the p-value will lead to a failure to reject the null hypothesis.

- **Comment** – The definition should refer less than or equal to. This point has no practical influence when the test statistic follows a continuous distribution as it is generally the case.
- **Example:** Resume example 12.14 using p-value.

Test (p-value): p-value= $\Pr(Z \geq z) = \Pr(\overline{X} \geq \overline{x} \mid \mu = 1200) = 0.3851$ do not reject H0 for $\alpha = 0.05$

Appendix 1 – Generalization of Example 12.4

A population has an exponential distribution with mean θ. We want to estimate the population mean using a sample of size $n = 2k + 1$. Determine the bias of the sample mean and the sample median as estimators of the population mean.

Sample mean:

$$E(\overline{X}) = \theta$$

Sample median

Let T be the sample median. Then

$$f_T(t) = \frac{(2k+1)!}{k!1!k!}\left(1 - e^{-t/\theta}\right)^k \left(e^{-t/\theta}\right)^k \theta^{-1} e^{-t/\theta} \qquad t > 0$$

$$f_T(t) = \frac{(2k+1)!}{(k!)^2}\left(e^{-(k+1)t/\theta}\right)\theta^{-1}\left(1-e^{-t/\theta}\right)^k \qquad t>0$$

$$= \frac{(2k+1)!}{(k!)^2}\left(e^{-(k+1)t/\theta}\right)\theta^{-1}\sum_{s=0}^{k}(-1)^s\,1^{k-s}\left(e^{-t/\theta}\right)^s\binom{k}{s}$$

$$= \frac{(2k+1)!}{(k!)^2}\left(e^{-(k+1)t/\theta}\right)\theta^{-1}\sum_{s=0}^{k}(-1)^s\,e^{-st/\theta}\binom{k}{s}$$

$$= \frac{(2k+1)!}{(k!)^2}\sum_{s=0}^{k}(-1)^s\,\frac{k+s+1}{\theta}e^{-(k+s+1)t/\theta}\,\frac{\binom{k}{s}}{k+s+1}$$

$$E(T \mid \theta) = \int_0^\infty t\, f_T(t)\, dt$$

$$= \int_0^\infty t \frac{(2k+1)!}{(k!)^2} \sum_{s=0}^{k} (-1)^s \frac{k+s+1}{\theta} e^{-(k+s+1)t/\theta} \frac{\binom{k}{s}}{k+s+1} dt$$

$$= \sum_{s=0}^{k} (-1)^s \frac{(2k+1)!}{(k!)^2} \frac{\binom{k}{s}}{k+s+1} \int_0^\infty t \frac{k+s+1}{\theta} e^{-(k+s+1)t/\theta} dt$$

$$= \sum_{s=0}^{k} (-1)^s \frac{(2k+1)!}{(k!)^2} \frac{\binom{k}{s}}{k+s+1} \frac{\theta}{k+s+1}$$

$$= \theta \sum_{s=0}^{k} (-1)^s \frac{(2k+1)!}{k!\,s!\,(k-s)!} \frac{1}{(k+s+1)^2}$$

That is

$$E(T \mid \theta) = \theta\, h(k) \quad \text{with} \quad h(k) = \sum_{s=0}^{k} (-1)^s \frac{(2k+1)!}{k!\,s!\,(k-s)!} \frac{1}{(k+s+1)^2}$$

As the median of the population is $\theta \ln 2$ we can imagine that, as the sample size increases, $h(k)$ will be closest to $\ln 2$. Note that the limit of $h(k)$ is not $\ln 2$ (I think).

Using Mathematica we can get values of $h(k)$ for different sample sizes. Remember that $\ln 2 \approx 0.693147$.

n	k	factor	n	k	factor
3	1	0.833333	41	20	0.705194
5	2	0.783333	61	30	0.701277
7	3	0.759524	101	50	0.698073
9	4	0.745635	201	100	0.695629
11	5	0.736544	1001	500	0.693646
21	10	0.71639	2001	1000	0.693397
31	15	0.709016	10001	5000	0.693646

Appendix 2 – Cramér Rao inequality

- Cramér-Rao Inequality applied to biased estimators
  Let $(X_1, X_2, \cdots, X_n)$ be a random sample from a population with probability density function $f_X(x \mid \theta)$ and let $T = T(X_1, X_2, \cdots, X_n)$ be any estimator of $\tau(\theta)$ satisfying $\frac{d}{d\theta} E(T \mid \theta) = \int_{D_{\mathbf{x}}} \frac{\partial}{\partial \theta} \big(T(\mathbf{x}) \times f(\mathbf{x} \mid \theta) \big) d\,\mathbf{x}$ and $\operatorname{var}(T \mid \theta) < \infty$.

  Then $\operatorname{var}(T \mid \theta) \geq \dfrac{\left(\frac{d}{d\theta} E(T \mid \theta) \right)^2}{n \Im(\theta)}$ where $\Im(\theta) = E\left(\frac{\partial}{\partial \theta} \ln f_X(X \mid \theta) \right)^2 = -E\left(\frac{\partial^2}{\partial \theta^2} \ln f_X(X \mid \theta) \right)$

  Proof: See Casella and Berger, 2nd edition, pages 335 to 337

Appendix 3 – Independence between $\overline{X}$ and S^2 when the population is normal

- Without loss of generality let us consider that our population has mean 0 and variance 1.
- Random sample $(X_1, X_2, \cdots, X_n)$. The density function of the random sample is given by

  $$f_{\mathbf{X}}(x_1, x_2, \cdots, x_n) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

- First step: Obtain $S^2 = (n-1)^{-1}\left(\left(\sum_{i=2}^{n}(X_i - \overline{X})\right)^2 + \sum_{i=2}^{n}(X_i - \overline{X})^2\right)$

  As $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ we get $(X_1 - \overline{X}) = (X_1 - \overline{X}) - \sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=2}^{n}(X_i - \overline{X})$

  then $\sum_{i=1}^{n}(X_i - \overline{X})^2 = (X_1 - \overline{X})^2 + \sum_{i=2}^{n}(X_i - \overline{X})^2 = \left(\sum_{i=2}^{n}(X_i - \overline{X})\right)^2 + \sum_{i=2}^{n}(X_i - \overline{X})^2$

  and finally $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = (n-1)^{-1}\left(\left(\sum_{i=2}^{n}(X_i - \overline{X})\right)^2 + \sum_{i=2}^{n}(X_i - \overline{X})^2\right)$

- Second step: As S^2 is a function of $(X_2-\overline{X}, X_3-\overline{X}, \cdots, X_n-\overline{X})$ we have to prove that $\overline{X}$ is independent of that vector. If we define $Y_1=\overline{X}$, $Y_2=X_2-\overline{X}$, $Y_3=X_3-\overline{X}$, …., $Y_n=X_n-\overline{X}$ we must prove that Y_1 is independent of $(Y_2,\cdots,Y_n)$.
- Third step: Define the joint distribution of $(Y_1,Y_2,\cdots,Y_n)$

  We have

$$\begin{cases} y_1=\overline{x} \\ y_2=x_2-\overline{x} \\ \cdots \\ y_n=x_n-\overline{x} \end{cases}$$

It is straightforward to see that $x_i=y_i+\overline{x}=y_i+y_1$ for $i=2,3,\cdots,n$ and that

$\overline{x}=y_1 \Leftrightarrow \sum_{i=1}^{n} x_i = n\,y_1 \Leftrightarrow x_1 = n\,y_1-\sum_{i=2}^{n} x_i = n\,y_1-\sum_{i=2}^{n} y_i-(n-1)\,y_1 = y_1-\sum_{i=2}^{n} y_i$ the inverse transformation is then

$$\begin{cases} x_1=y_1-\sum_{i=2}^{n} y_i \\ x_2=y_2+y_1 \\ \cdots \\ x_n=y_n+y_1 \end{cases}$$ and the Jacobian will be $J=\begin{vmatrix} 1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & \cdots & 1 \end{vmatrix}=n$

Then

$$
\begin{aligned}
f_{\mathbf{Y}}(y_1, y_2, \cdots, y_n) &= (2\pi)^{-n/2}\, n\, \exp\left(-\frac{1}{2}\left(\left(y_1 - \sum\nolimits_{i=2}^{n} y_i\right)^2 + \sum\nolimits_{i=2}^{n} (y_i + y_1)^2\right)\right) \\
&= (2\pi)^{-n/2}\, n\, \exp\left(-\frac{1}{2}\left(y_1^2 - 2\, y_1 \sum\nolimits_{i=2}^{n} y_i + \left(\sum\nolimits_{i=2}^{n} y_i\right)^2 + \sum\nolimits_{i=2}^{n} y_i^2 + \sum\nolimits_{i=2}^{n} y_1^2 + 2\, y_1 \sum\nolimits_{i=2}^{n} y_i\right)\right) \\
&= (2\pi)^{-n/2}\, n\, \exp\left(-\frac{1}{2}\left(y_1^2 + \left(\sum\nolimits_{i=2}^{n} y_i\right)^2 + \sum\nolimits_{i=2}^{n} y_i^2 + (n-1)\, y_1^2\right)\right) \\
&= (2\pi)^{-n/2}\, n\, \exp\left(-\frac{1}{2}\left(n\, y_1^2 + \left(\left(\sum\nolimits_{i=2}^{n} y_i\right)^2 + \sum\nolimits_{i=2}^{n} y_i^2\right)\right)\right) \\
&= (2\pi)^{-1/2}\, n^{1/2} \exp\left(-\frac{n\, y_1^2}{2}\right) \times (2\pi)^{-(n-1)/2}\, n^{1/2} \exp\left(-\frac{1}{2}\left(\left(\sum\nolimits_{i=2}^{n} y_i\right)^2 + \sum\nolimits_{i=2}^{n} y_i^2\right)\right)
\end{aligned}
$$

As the joint distribution is the product of the marginal distributions, Y_1 is independent of $(Y_2, \cdots, Y_n)$.

Appendix 4 – Confidence interval for Bernoulli populations (large samples)

$$Q(X_1, X_2, \ldots, X_n, \theta) = \frac{\bar{X} - \theta}{\sqrt{\mathrm{var}(\bar{X})}} = \frac{\bar{X} - \theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \overset{\circ}{\sim} N(0,1)$$

$$\Pr\left(-z_{\alpha/2} \le Q(X_1, X_2, \ldots, X_n, \theta) \le z_{\alpha/2}\right) = \Pr\left(-z_{\alpha/2} \le \frac{\bar{X} - \theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \le z_{\alpha/2}\right) = \Pr\left(\left|\frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n}\right| \le z_{\alpha/2}\right)$$

$$\left|\frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n}\right| \le z_{\alpha/2} \Leftrightarrow \left(\frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}}\sqrt{n}\right)^2 \le z_{\alpha/2}^2 \Leftrightarrow \frac{(\bar{X} - \theta)^2}{\theta(1-\theta)} \le \frac{z_{\alpha/2}^2}{n} \Leftrightarrow (\bar{X} - \theta)^2 - \frac{z_{\alpha/2}^2}{n}\theta(1-\theta) \le 0$$

$$\Leftrightarrow \bar{X}^2 + \theta^2 - 2\theta\bar{X} - \frac{z_{\alpha/2}^2}{n}\theta + \frac{z_{\alpha/2}^2}{n}\theta^2 \le 0$$

$$\Leftrightarrow \theta^2\left(1 + \frac{z_{\alpha/2}^2}{n}\right) - \theta\left(2\bar{X} + \frac{z_{\alpha/2}^2}{n}\right) + \bar{X}^2 \le 0$$

Then, the possible values of θ have to be between the 2 roots of the equation.

$$\Delta = \left(2\overline{X} + \frac{z_{\alpha/2}^2}{n}\right)^2 - 4\left(1 + \frac{z_{\alpha/2}^2}{n}\right)\overline{X}^2 = 4\overline{X}^2 + \left(\frac{z_{\alpha/2}^2}{n}\right)^2 + 4\overline{X}\left(\frac{z_{\alpha/2}^2}{n}\right) - 4\overline{X}^2 - 4\left(\frac{z_{\alpha/2}^2}{n}\right)\overline{X}^2$$

$$= \left(\frac{z_{\alpha/2}^2}{n}\right)^2 + 4\left(\frac{z_{\alpha/2}^2}{n}\right)\overline{X} - 4\left(\frac{z_{\alpha/2}^2}{n}\right)\overline{X}^2 = \left(\frac{z_{\alpha/2}^2}{n}\right)^2 + 4\left(\frac{z_{\alpha/2}^2}{n}\right)\overline{X}(1-\overline{X}) = 4z_{\alpha/2}^2\left(\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}\right)$$

$$\theta_1 = \frac{\left(2\overline{X} - \frac{z_{\alpha/2}^2}{n}\right) - \sqrt{4z_{\alpha/2}^2\left(\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}\right)}}{2\left(1 + \frac{z_{\alpha/2}^2}{n}\right)} = \frac{\overline{X} - \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}}}{\left(1 + \frac{z_{\alpha/2}^2}{n}\right)}$$

$$\theta_2 = \frac{\overline{X} - \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\overline{X}(1-\overline{X})}{n}}}{\left(1 + \frac{z_{\alpha/2}^2}{n}\right)}$$

Appendix 5 – Unbiased estimator for σ (normal populations)

Let $(X_1, X_2,..., X_n)$ be a random sample of size n from a normal population with mean μ and variance σ^2.

As it is well known, $Q = \dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ with $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}$.

Result 1 – If $Y \sim \chi^2_{(r)}$ then $E(\sqrt{Y}) = \dfrac{\Gamma\left((r+1)/2\right)\sqrt{2}}{\Gamma\left(r/2\right)}$

Proof: $f_Y(y) = \dfrac{1}{\Gamma\left(r/2\right)2^{r/2}}\, y^{r/2-1}\, e^{-y/2}$

$$E(\sqrt{Y}) = \int_0^\infty y^{1/2} f_Y(y)\,dy = \int_0^\infty y^{1/2} \frac{1}{\Gamma\left(r/2\right)2^{r/2}}\, y^{r/2-1}\, e^{-y/2}\,dy = \frac{1}{\Gamma\left(r/2\right)2^{r/2}} \int_0^\infty y^{(r+1)/2-1}\, e^{-y/2}\,dy$$

$$= \frac{\Gamma\left((r+1)/2\right)2^{(r+1)/2}}{\Gamma\left(r/2\right)2^{r/2}} \int_0^\infty \frac{1}{\Gamma\left((r+1)/2\right)2^{(r+1)/2}}\, y^{(r+1)/2-1}\, e^{-y/2}\,dy$$

$$= \frac{\Gamma\left((r+1)/2\right)2^{(r+1)/2}}{\Gamma\left(r/2\right)2^{r/2}}$$

The value of the integral is 1 (density of a $\chi^2_{(r+1)}$ over its domain)

$$= \frac{\Gamma\left((r+1)/2\right)\sqrt{2}}{\Gamma\left(r/2\right)}$$

Applying result 1 to *Q*, we get $E(\sqrt{Q})=\dfrac{\Gamma(n/2)\sqrt{2}}{\Gamma((n-1)/2)}$ but $E(\sqrt{Q})=E\left(\sqrt{\dfrac{(n-1)\,S^2}{\sigma^2}}\right)=\sqrt{\dfrac{(n-1)}{\sigma^2}}E(S)$. Then

$$\sqrt{\frac{(n-1)}{\sigma^2}}E(S)=\frac{\Gamma(n/2)\sqrt{2}}{\Gamma((n-1)/2)}\Leftrightarrow E(S)=\frac{\Gamma(n/2)\sqrt{2}}{\Gamma((n-1)/2)\sqrt{n-1}}\sigma$$

Consequently an unbiased estimator will be given by

$$T=\frac{\Gamma((n-1)/2)\sqrt{n-1}}{\Gamma(n/2)\sqrt{2}}S\,.$$