ISEG - Lisbon School of Economics and Management
Statistics I
(Date of this version: 2/5/2018)
Handout 6 – Special random variables

Issues covered:

6.1 Discrete Random Variables

6.1.1 The discrete uniform random variable

A random variable X has a *discrete uniform distribution* and it is referred to as a discrete uniform random variable if and only if , its probability function is given by

$$f_X(x_j) = \frac{1}{k},\ j = 1, 2, 3, ..., k$$

where $x_j \neq x_i$ for $i \neq j \quad D_X = \{x_1, x_2, ..., x_k\}$

Properties:

1. $\mu_X = E(X) = \sum_{i=1}^{k} x_i/k$
2. $Var(X) = \sum_{i=1}^{k} x_i^2/k - \left(\sum_{i=1}^{k} x_i/k\right)^2$
3. $M_X(t) = \sum_{i=1}^{k} e^{tx_i}/k$

Example: The throwing a fair dice and X is the the number of dots showing on its upper surface. The possible values of X are 1, 2, 3, 4, 5, 6 with $P(X = i) = 1/6$, $i = 1, 2, 3, 4, 5, 6$.

6.1.2 The Bernoulli random variable

The *Bernoulli random variable* [named after the Swiss mathematician Jacob Bernoulli (1654-1705)] takes the value 1 with probability p and the value 0 with probability $1-p$, where $p \in (0, 1)$, that is

$$X = \begin{cases} 1 & \text{where } P(X = 1) = p \\ 0 & \text{where } P(X = 0) = 1 - p \end{cases}$$

the probability function is given by

$$f_X(x) = \mathcal{P}(X = x) = p^x (1-p)^{1-x}, x = 0, 1.$$

Properties:

1. $E(X) = p$
2. $Var(X) = p(1-p)$
3. $M_X(t) = (1-p) + pe^t$.

6.1.3 The Binomial random variable

The *Binomial random variable* is defined as the number of successes in n trials, each of which has the probability of success p.

Remark: If $n = 1$ the Binomial random variable corresponds to the Bernoulli random variable.
Example 1: Suppose $n = 2$, for instance $X =$ number of boys in a family of 2 children.
Let us calculate the probability of $0, 1, 2$ boys in 2 births and define $P(boy) = p$
We can have 4 possible cases:

$$(boy, boy)\,;(boy, girl)\,;(girl, boy)\,;(girl, girl)$$

Hence:

- $P(X = 0) = P\,(girl, girl) = (1-p)^2$
- $P(X = 1) = P((boy, girl) \text{ or } (girl, boy)) = P\,(boy, girl) + P\,(girl, boy) = 2p(1-p)$
- $P(X = 2) = P\,(boy, boy) = p^2$.

Example 2: Suppose $n = 3$, for instance $X =$ number of boys in a family of 3 children.
Let us calculate the probability of $0, 1, 2, 3$ boys in 3 births.
We can have 8 possible cases:

$$\begin{aligned}&(boy, boy, boy)\,;(boy, girl, boy)\,;(girl, boy, boy)\,;(girl, girl, boy)\,;\\&(boy, boy, girl)\,;(boy, girl, girl)\,;(girl, boy, girl)\,;(girl, girl, girl)\,.\end{aligned}$$

Hence:

- $P(X = 0) = P\,(girl, girl, girl) = (1-p)^3$.
- $P(X = 1) = P((girl, girl, boy) \text{ or } (boy, girl, girl) \text{ or } (girl, boy, girl)) = 3(1-p)^2 p$.
- $P(X = 2) = P((boy, girl, boy) \text{ or } (girl, boy, boy) \text{ or } (boy, boy, girl)) = 3(1-p)p^2$.
- $P(X = 3) = P\,(boy, boy, boy) = p^3$.

The Binomial random variable: $X =$ number of successes in n trials. One can show that the probability function is given by

$$f_X(x) = \binom{n}{x} \times p^x (1-p)^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is the number of x combinations from a set with n elements and $k! = k \times (k-1) \times ... \times 2 \times 1$

Exercise: What is the probability of the number of boys is equal to 3 in a family of 6 children when $P(boy) = p = 0.5$? What is the the probability of the number of boys is less or equal to 3?

Remark:

- The parameters of the random variable are n and p.
- If X is a Binomial random variable with parameters n and p we write $X \sim B(n, p)$.
- In the case of the Bernoulli random variable $X \sim B(1, p)$.

Properties:

1. $E(X) = np$,
2. $Var(X) = np(1 - p)$
3. $M_X(t) = \left[(1 - p) + pe^t\right]^n$
4. If $X_i \sim B(1, p)$ and the X_i are independent random variables $\sum_{i=1}^{n} X_i \sim B(n, p)$, that is the sum of n independent Bernoulli random variables with parameter p is a Binomial random variable with parameters r and p.
5. If $X_1 \sim B(n_1, p)$ and $X_2 \sim B(n_2, p)$ and X_1 and X_2 are independent, then $X_1 + X_2 \sim B(n_1 + n_2, p)$

6.1.4 The Poisson random variable

The Poisson random variable, named after the French mathematician Simeon-Denis Poisson (1781-1840), is applicable in many situations where rare events occur.

The Poisson random variable describes the number of occurrences within a randomly chosen unit of time or space. For example, within a minute, hour, day, kilometer.

Examples:

- in the inspection and quality control of manufactured goods the number of defective articles in a large lot can be expected to be small.
- number of customers arriving at a cash point in a given minute.
- number of file server virus infections at a data center during a 24-hour period.

Famous example: Bortkiewiz in 1898 used this distribution to study the number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry.

The Poisson distribution's only parameter is λ: λ represents the mean number of events per unit of time or space.

The Poisson probability function is a discrete function defined for non-negative integers. The Poisson distribution with parameter $\lambda > 0$, it is defined by

$$f_X(x) = P\left(X = x\right) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ..$$

Remark: If X is a Poisson random variable with parameter λ, we write $X \sim Poisson(\lambda)$.

Properties:

1. $E(X) = \lambda$.
2. $Var(X) = \lambda$.
3. $M_X(X) = e^{\lambda(e^t - 1)}$.
4. If $X_i \sim Poisson(\lambda_i)$ and the X_i are independent random variables $\sum_{i=1}^{n} X_i \sim Poisson\left(\sum_{i=1}^{n} \lambda_i\right)$, that is the sum of n independent Poisson random variables with parameter λ_i is a Poisson random variable with parameter $\sum_{i=1}^{n} \lambda_i$.

Exercise: On Thursday morning between 9 A.M. and 10 A.M. customers arrive and enter the queue at a bank branch with mean rate of 1.7 customers per minute. Assuming the the number of customers is a Poisson random variable:

What is the probability that two or fewer customers will arrive in a given minute?

What is the probability of at least three customers (the complimentary event)?

Theorem: (The law of rare events). If Y is a Binomial random variable with parameters n and $p = \lambda/n$ then,

$$\begin{aligned} \lim_{n\to\infty} f_Y(y) &= \lim_{n\to\infty} \binom{n}{y} \times p^y (1-p)^{n-y} \\ &= \frac{\lambda^y e^{-\lambda}}{y!}, \end{aligned}$$

That is the limit of the probability function of the binomial random variable with parameters n and $p = \lambda/n$ is the Poisson random variable with parameter λ.

Remark: Hence the Poisson distribution can be used to approximate the Binomial distribution when the number of trials n is large and the probability of success p is small (note that since n is large $p = \lambda/n$ is small).

Exercise: A corporation has 250 personal computers. The probability that any one of them will require repair in a given week is 0.01. Find the probability that fewer than 4 of the personal computers will require repair in a given week.

6.2 Continuous random variables

6.2.1 The continuous uniform *random variable*

The probability density function of the *uniform random variable* on an interval (a, b), where $a < b$, is the function

$$f_X(x) = \begin{cases} 0 & if \quad x \le a \\ \frac{1}{b-a} & if \quad a < x < b \\ 0 & if \quad b \le x \end{cases}$$

The cumulative distribution function is the function

$$F_X(x) = \begin{cases} 0 & if \quad x \le a \\ \frac{x-a}{b-a} & if \quad a \le x \le b \\ 1 & if \quad b \le x \end{cases}$$

Remark: If X is a *uniform random variable* in the interval (a, b) we write $X \sim U(a, b)$.

Properties:

1. The moment generating function

$$M_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & if \quad t \neq 0 \\ 1 & if \quad t = 0. \end{cases}$$

(The moment-generating function is not differentiable at zero, but the moments can be calculated by differentiating and then taking $\lim_{t\to 0}$)

2. Moments about the origin

$$E(X^k) = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)}, k = 1, 2, 3, ...$$

3. $E(X) = (a+b)/2$.
4. $Var(X) = (b-a)^2/12$.
5. $Skewness = \gamma_1 = 0$.

Theorem: (Probability Integral Transformation) Let X be a random variable with a strictly increasing cumulative distribution function $F_X(x)$, then $Y = F_X(X) \sim U(0,1)$. Conversely, if $Y \sim U(0,1)$, then $X = F_X^{-1}(Y)$ is a continuous random variable with cumulative distribution function $F_X(x)$.

Remark: This theorem is very useful in simulation problems.

6.2.2 Exponential Random variable

The probability density function of an *exponential random variable* with parameter λ is

$$f_X(x) = \begin{cases} 0 & if \quad x < 0 \\ \lambda e^{-\lambda x} & if \quad x \geq 0 \end{cases}$$

1. The cumulative distribution function is given by

$$F_X(x) = \begin{cases} 0 & if \quad x < 0 \\ 1 - e^{-\lambda x} & if \quad x \geq 0 \end{cases}$$

The exponential probability distribution may be used for random variables such as the time between arrivals at a car wash, the time required to load a truck, the distance between major defects in a highway, and so on.

Remark: If X is an exponential random variable with parameter λ we write $X \sim Exp(\lambda)$.

Properties:

1. Moment Generating Function $M_X(t) = (1 - t/\lambda)^{-1}$ $t < \lambda$.
2. $E(X) = 1/\lambda$.
3. $Var(X) = 1/\lambda^2$.
4. Lack of memory: $P(X > x + s | X > x) = P(X > s)$ for any $x \geq 0$ and $s \geq 0$.

5. Let $X_i \sim Exp(\lambda_i)$, $i = 1, 2, ..., k$, be independent random variables, then $Y = \min\{X_1, X_2, ..., X_k\} \sim Exp(\sum_{i=1}^{k} \lambda_i)$

Remark: Property 4 means the following. Suppose that X is the waiting time until the arrival of a customer in a shop, $x = 60$ minutes and $a = 15$ minutes. The lack of memory property means that the probability of waiting more than 75 minutes for a customer given that you already waited more than 60 minutes, is equal to the probability of waiting more than 15 minutes for the customer. Hence, the lack of memory property means that it does not matter how long you have waited so far.

Exponential random variables (sometimes) give good models for the time to failure of mechanical devices. For example, we might measure the number of kilometers traveled by a given car before its transmission ceases to function. Suppose that this distribution is governed by the exponential distribution with mean 100000. What is the probability that a car's transmission will fail during its first 50000 kilometers of operation?

6.2.3 The Normal random variable

The most famous continuous distribution is the *normal distribution* (introduced by Abraham de Moivre, 1667-1754). The normal probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The cumulative distribution function does not have a close form solution:

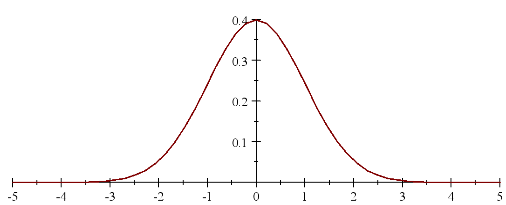
$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

When a random variable X has the is normal with parameters μ and σ^2 we write $X \sim N(\mu, \sigma^2)$.
Properties:

1. Moment generating function $M_X(t) = e^{\left(\mu t + 0.5\sigma^2 t^2\right)}$
2. $E(X) = \mu$.
3. $Var(X) = \sigma^2$
4. $skewness = \gamma_1 = 0$.
5. $kurtosis = \gamma_2 = 3$

Remark: The excess kurtosis of any random variable is defined as $\gamma_2^* = \gamma_2 - 3$. Hence for the normal random variable $\gamma_2^* = 0$.
Remarks:

- When $\mu = 0$ and $\sigma^2 = 1$, the distribution is denoted as *redstandard normal.* Its shape is the following:



- The cumulative distribution function of Z is tabulated.

The probability density function of the standard normal distribution is denoted $\phi(z)$ and it is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

The standard normal cumulative distribution function is denoted as

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)\, dt.$$

Properties of the standard normal cumulative distribution function:

- $P(Z > z) = 1 - \Phi(z)$.
- $P(Z < -z) = P(Z > z)$.
- $P(|Z| > z) = 2[1 - \Phi(z)]$, for $z > 0$.

Theorem: (Linear combinations of Normal random variables): Let X and Y be two independent random variables such that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Let $V = aX + bY + c$, then

$$V \sim N(\mu_V, \sigma_V^2)$$

where

$$\begin{aligned} \mu_V &= a\mu_X + b\mu_Y + c \\ \sigma_V^2 &= a^2\sigma_X^2 + b^2\sigma_Y^2. \end{aligned}$$

Remarks:

- A special case is obtained when $b = 0$, if $V = aX + c$, then $V \sim N(\mu_V, \sigma_V^2)$ where $\mu_V = a\mu_X + c$, $\sigma_V^2 = a^2\sigma_X^2$.
- if $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Exercise: Compute $P(4 < X < 7)$, where $X \sim N(5, 2)$.

Theorem: If the random variable $X_i, i = 1, ..., n$ have a normal distribution, $X_i \sim N(\mu_i, \sigma_i^2)$, and are independent, then

$$\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2).$$

Exercise: At an establishment that sells building materials, it is known that daily sales of sand (in kgs) have a random behavior, translated by a Normal distribution with mean 20 and standard deviation 20. Assuming independence of the daily sales in a month, what is the probability that in any given month (20 days) sales exceed half ton of sand?

6.2.4 The gamma and the chi-squared random variables

The *gamma cumulative distribution* function is defined for $x > 0$, $a > 0$, $b > 0$, by the integral

$$F_X(x) = \frac{1}{b^a \Gamma(a)} \int_0^x u^{a-1} e^{-\frac{u}{b}} du$$

where $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$ is the Gamma function. The parameters a and b are called the shape parameter and scale parameter, respectively.

The probability density function for the gamma distribution is

$$f_X(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

Remarks:

1. If X is a gamma random variable with parameters a and b we write $X \sim Gamma(a, b)$
2. The gamma function satisfies

$$\Gamma(1) = 1 \text{ and } \Gamma(t+1) = t\Gamma(t)$$

   and for positive integers k, it is the familiar factorial function

$$\Gamma(k) = (k-1)!$$

3. if $a = 1$ and $\frac{1}{b} = \lambda$, $X \sim Exp(\lambda)$, that is $Exp(\lambda) = Gamma(1, \frac{1}{\lambda})$.
4. **Important case:** When $a = v/2$ and $b = 2$ we have the chi-squared distribution which has the notation $\chi^2(v)$, that is $\chi^2(v) = Gamma(v/2, 2)$. v is known as degrees of freedom (in the Tables $v = df$).

Properties

1. The Moment generating function of the Gamma distribution is given by: $M_X(t) = (1 - bt)^{-a}$ for $t < 1/b$
2. $E(X^k) = \frac{b^k \Gamma(a+k)}{\Gamma(a)}$, $k = 1, 2, 3, ...$
3. $E(X) = ab$.
4. $Var(X) = ab^2$.
5. Let X_1, X_2, be independent random variables with Gamma distribution $X_1 \sim Gamma(a_1, b)$ and $X_2 \sim Gamma(a_2, b)$, then $X_1 + X_2 \sim Gamma(a_1 + a_2, b)$.
6. Let X_1, X_2,..., X_n be independent random variables with Gamma distribution $X_i \sim Gamma(a_i, b)$, $i = 1, ..., n$, then $\sum_{i=1}^n X_i \sim Gamma(\sum_{i=1}^n a_i, b)$.
7. If $X \sim Gamma(a, b)$, then $2X/b \sim \chi^2(2a)$

Remarks:

1. Property 7 is important because while there are no tables of the gamma distribution, there are tables of the chi-squared distribution.

2. In the tables we find the value c_α such that $P(X > c_\alpha) = \alpha$, where $X \sim \chi^2(v)$.

3. If v is very large ($v > 100$) we should use the result:

$$X \sim \chi^2(v) \Rightarrow \frac{X - v}{\sqrt{2v}} \overset{a}{\sim} N(0,1) \text{ as } v \to \infty$$

where $\overset{a}{\sim}$ means that it is asymptotic distribution (n large).

Exercise: Let X be the compensation paid by an insurer for certain risk. Assume that $X \sim Gamma(2, 125)$. Compute $P(X > 210)$.

In the case of the chi-squared random variables we have:

1. $E(X) = v$.

2. $Var(X) = 2v$.

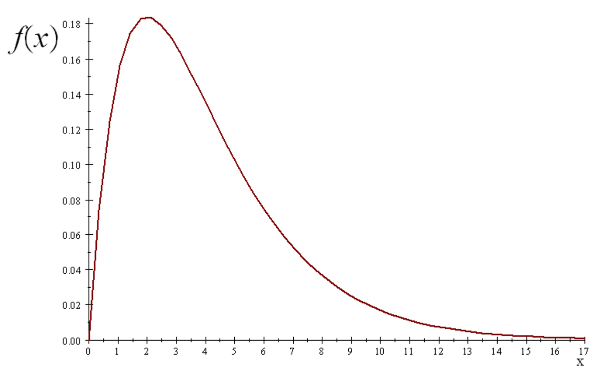
3. Let X_1, X_2, be independent random variables with Chi-squared distribution $X_1 \sim \chi^2(v_1)$ and $X_2 \sim \chi^2(v_2)$, then $X_1 + X_2 \sim \chi^2(v_1 + v_2)$.

4. Let X_1, X_2,...,X_k be independent random variables with Chi-squared distribution $X_1 \sim \chi^2(v_1)$ and $X_2 \sim \chi^2(v_2)$, ..., $X_k \sim \chi^2(v_k)$, then $\sum_{i=1}^{k} X_i \sim \chi^2\left(\sum_{i=1}^{k} v_i\right)$.

5. If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$.

6. Combining properties 4 and 5: Let Z_i, $i = 1, ..., q$ be independent random variables each distributed as standard normal. Define $X = \sum_{i=1}^{q} Z_i^2$. Then $X \sim \chi^2(q)$.

Probability Density function of the Chi-Square Distribution with 4 *df*



$q = 4$

6. The Central Limit Theorem

6.3 The Central Limit Theorem

If the random variables $X_i, i = 1, ..., n$ have a normal distribution, $X_i \sim N(\mu_i, \sigma_i^2)$, and are independent, then

$$\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2).$$

- Assuming that $\mu_i = \mu_X$ and $\sigma_i^2 = \sigma_X^2$, for $i = 1, ..., n$ we have

  $$\sum_{i=1}^{n} X_i \sim N(n\mu_X, n\sigma_X^2).$$

  Thus

  $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu_X, \sigma_X^2/n).$$

  If we standardize we have

  $$Z = \frac{\overline{X} - \mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$$

However, what happens if the $X_i's$ are not normally distributed?
The answer is given by the Central Limit Theorem:

Theorem: (*The Central Limit Theorem* - Lindberg-Levy) If the X_i, $i = 1, ..., n$ are independent, and $E(X_i) = \mu_X$ and variance $Var(X_i) = \sigma_X < +\infty$, then the distribution of

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X}$$

converges to a standard normal distribution as n tends to infinity. We write $Z \overset{a}{\sim} N(0,1)$ where the symbol $\overset{a}{\sim}$ reads "distributed asymptotically" (it means that if the sample size is large the distribution of Z is close to the standard normal).

Remarks:

1. The Central Limit Theorem is valid for discrete or continuous random variables
2. When is n large enough? Depends on the distribution of X.
   (a) For unimodal symmetric distributions convergence is faster and the approximation better.
   (b) It is advised to use the CLT only for $n \geq 30$.

Example: A large freight elevator can transport a maximum of 4400 kilos. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean 93 kilos and standard deviation 6.8 kilos. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

A special case of the Central Limit Theorem of Lindberg-Levy is the Central Limit Theorem of De Moivre-Laplace, which corresponds to the case that each X_i is Bernoulli with parameter $p = P(X_i = 1)$.

Theorem: (The Central Limit Theorem - De Moivre-Laplace) If the X_i, $i = 1, ..., n$ are independent Bernoulli random variables with $p = P(X_i = 1) \in (0, 1)$ then

$$Z = \frac{\sqrt{n}\left(\overline{X} - p\right)}{\sqrt{p(1-p)}}$$

converges to a standard normal distribution as n tends to infinity. We write $Z \overset{a}{\sim} N(0, 1)$.
Remarks:

- $\bar{X}$ in this case is a proportion, that is $\bar{X} = s/n$, where s is the number of successes.
- Note that now $\mu_X = p$ and $\sigma_X = \sqrt{p(1-p)}$
- Note that

$$\begin{aligned} Z &= \frac{\sqrt{n}\left(\overline{X} - p\right)}{\sqrt{p(1-p)}} \\ &= \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \end{aligned}$$

  and $\sum_{i=1}^{n} X_i \sim B(n, p)$.

Remarks:

1. If we want to compute $P\left(a \leq \sum_{i=1}^{n} X_i \leq b\right)$ we have

$$\begin{aligned} P\left(a \leq \sum_{i=1}^{n} X_i \leq b\right) &= P\left(\frac{a - n\mu_X}{\sqrt{n}\sigma_X} \leq \frac{\sum_{i=1}^{n} X_i - n\mu_X}{\sqrt{n}\sigma_X} \leq \frac{b - n\mu_X}{\sqrt{n}\sigma_X}\right) \\ &\simeq \Phi\left(\frac{b - n\mu_X}{\sqrt{n}\sigma_X}\right) - \Phi\left(\frac{a - n\mu_X}{\sqrt{n}\sigma_X}\right) \end{aligned}$$

   using the central limit theorem.

2. Note that if the X_i are discrete random variables is follows that $\sum_{i=1}^{n} X_i$ is also discrete, and the above approximation is poor. In this case it is advisable to use the continuity correction

$$P\left(a \leq \sum_{i=1}^{n} X_i \leq b\right) \simeq \Phi\left(\frac{b + 0.5 - n\mu_X}{\sqrt{n}\sigma_X}\right) - \Phi\left(\frac{a - 0.5 - n\mu_X}{\sqrt{n}\sigma_X}\right).$$

Exercise: Suppose a fair coin is tossed 200 times and let Y be the number of "heads" in these 200 tosses. Compute $P(95 \leq Y \leq 105)$.