

# stat240 notes

Fall 2021

# 1 Counting

**1.1 Theorem.** Number of subsets of size  $k$  The number of subsets of size  $k$  selected from a set of size  $n$  are:

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

**1.2 Theorem.** Counting with repeated symbols/objects If we have  $n_i$  symbols of type  $i$  with  $i = 1, 2, \dots, k$  with

$$n_1 + n_2 + \dots + n_k = n$$

Then the number of arrangements using all the symbols is:

$$\begin{aligned} & \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n_k}{n_k} \\ &= \frac{n!}{n_1! n_2! \dots n_k!} \end{aligned}$$

**1.3 Theorem.** Hyper geometric identity

$$\sum_{x=0}^{\infty} \binom{a}{x} \binom{b}{n-x} = \binom{a+b}{n}$$

*Proof.*

$$\begin{aligned} (1+y)^{a+b} &= (1+y)^a (1+y)^b \\ &= \sum_{k=0}^{a+b} \binom{a+b}{k} y^k = \sum_{i=0}^a \binom{a}{i} y^i \sum_{j=0}^b \binom{b}{j} y^j \end{aligned}$$

The coefficient of  $y^k$  on the left hand side is  $\binom{a+b}{k}$  and the coefficient of  $y^k$  on the right hand side is

$$\sum_{i=0}^a \binom{a}{i} \binom{b}{k-i}$$

both coefficients must be equal. □

**1.4 Theorem.** Exponential series

$$e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}$$

## 2 Probability rules

### 2.1 Theorem. Set properties

- (1) Distribution law:

$$A(B \cup C) = AB \cup AC$$

- (2) Morgan's Laws:

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

### 2.2 Definition. Probability rules

- (1)  $P(\mathcal{S}) = 1$  where  $\mathcal{S}$  is the full sample space.
- (2) For any event  $A$ ,  $0 \leq P(A) \leq 1$ .
- (3) If  $A$  and  $B$  are events with  $A \subseteq B$  then  $P(A) \leq P(B)$
- (4) Addition Law:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

- (5) Complement events

$$P(A) = 1 - P(A^c)$$

### 2.3 Definition. Independent events 2 $A, B$ events are independent $\iff P(AB) = P(A)P(B)$

### 3 Conditional Probability

**3.1 Definition.** Probability  $A$  occurs given  $B$  has occurred

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

**3.2 Theorem. Multiplication Theorem** For  $n$  events  $A_1, A_2, \dots, A_n$  we have

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 A_1)P(A_4|A_3 A_2 A_1) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$$

**3.3 Theorem. Total probability** For any event  $A$  we have

$$P(A) = P(AB) + P(AB^c)$$

In general if we have a partition  $A_1, A_2, \dots$ , of  $\mathcal{S}$ , then

$$P(B) = P(B\mathcal{S}) = \sum_{i=1}^{\infty} P(A_i B) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i)$$

**3.4 Theorem.** Bayes formula

$$P(A_i|B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)}$$

When  $A_1, A_2, \dots$  is a partition of  $\mathcal{S}$  and  $P(B) \neq 0$ .

## 4 Discrete Random Variables

**4.1 Definition.** A random variable  $X$  over  $\mathcal{S}$  is a function

$$X : \mathcal{S} \rightarrow \mathbb{R}$$

$$X = X(\omega), \omega \in \mathcal{S}$$

**4.2 Definition.** Probability mass function

$$f(x_i) = P[X = x_i], i = 1, 2, \dots$$

Some properties of *pmf*

$$\sum_{\text{all } x} f(x) = 1$$

**4.3 Definition.** cumulative distribution function

$$F(x) = P[X \leq x] \quad \forall x \in \mathbb{R}$$

Where

$$(X \leq x) = \{\omega : X(\omega) \leq x, \omega \in \mathcal{S}\}$$

Some properties of *cdf*:

- (1) Non-decreasing  $x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$
- (2) Bounded  $0 \leq F(x) \leq 1$
- (3) Limits using the monotone convergence theorem we have

$$\lim_{x \rightarrow \infty} F(x) = 1$$

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

We can calculate  $f(x)$  using  $F(x)$  if  $X$  takes integer values then

$$f(x) = P[X = x] = P[X \leq x] - P[X \leq x - 1] = F(x) - F(x - 1)$$

$$F(x) = \sum_{u \leq x} f(u)$$

## 4.1 Bernoulli trials and related random variables

**4.4 Definition.** Bernoulli trials Experiments with only 2 outcomes Success/Failure. Let  $B = \text{Success}$  with  $P(B) = p$ . Then  $B^c = \text{Failure}$  with  $P(B^c) = 1 - p$ .

**4.5 Definition.** Bernoulli Random Variable Let  $X = X(\omega) = 1$  if  $\omega = B$  otherwise  $X(\omega) = 0$  if  $\omega = B^c$ . So the *pmf* of  $X$  is

$$\begin{aligned} f(1) &= p \\ f(0) &= 1 - p \end{aligned}$$

The distribution is

$$f(x) = p^x(1 - p)^{1-x}$$

Where  $x \in \{0, 1\}$

## 4.2 Binomial distribution

If we repeat independent Bernoulli trials with the same probability of success  $p$ , each  $n$  times then let  $X$  be the number of successes among  $n$  trials.

**4.6 Theorem.** Binomial distribution If  $X \sim \text{Binomial}(n, p)$  then the *pmf* is given by

$$f(x) = P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}$$

### 4.3 Hypergeometric Distribution

If we have a collection of  $N$  objects which can be classified into 2 types success ( $S$ ) and failure. If there are  $r$  successes and  $N - r$  failures. Pick  $n$  objects at random **without replacement**. Let  $X$  be the number of successes obtained,  $X$  has a hyper-geometric distribution.

$$X \sim \text{Hypergeometric}(N, r, n)$$

**4.7 Theorem.** hyper-geometric distribution If  $X \sim \text{Hypergeometric}(N, r, n)$  then the *pmf* is given by

$$f(x) = P[X = x] = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

*Proof.*  $\#\mathcal{S} = \binom{N}{n}$  total ways of choosing  $n$  objects. First we choose  $x$  successes from the total  $r$  and then choose the rest  $n - x$  from the  $N - r$  failures. Using the product rule we get the *pmf*.  $\square$



## 4.4 Negative Binomial Distribution

The experiment has two distinct outcomes  $S$  and  $F$  which is repeated independently with  $P(S) = p$ . Continue doing the experiment until we get  $k$  successes. Then let  $X$  be the number of failures before the  $k$ -th success is obtained then  $X$  has a Negative Binomial Distribution.

$$X \sim \text{NBinomial}(k, p)$$

There will be  $x + k$  trials where we get  $x$  failures and  $k$  successes then we stop. The last trial must be a success. So in the first  $x + k - 1$  trials we need  $x$  failures and  $k - 1$  successes.

The number of ways to get that is  $\binom{x+k-1}{x}$  we choose the  $x$  failures and the rest must be successes.

**4.8 Theorem.** Negative Binomial Distribution If  $X \sim \text{NBinomial}(k, p)$  then the *pmf* is given by

$$f(x) = P[X = x] = \binom{x+k-1}{x} p^k (1-p)^x$$

## 4.5 Geometric distribution

Consider the Negative Binomial Distribution with  $k = 1$ . That is we only need 1 success. So we repeat the trials until we get the first success.

$$X \sim \text{Geo}(p)$$

**4.9 Theorem.** Geometric Distribution If  $X \sim \text{Geo}(p)$  then the *pmf* is given by

$$f(x) = P[X = x] = p(1-p)^x$$

*Proof.*  $k = 1$  in Negative Binomial Distribution. □

## 4.6 Poisson distribution

### 4.6.1 Limit of Binomial

The Poisson distribution happens as a limiting case of the binomial distribution. That is  $n \rightarrow \infty$  and  $p \rightarrow 0$ . We can keep  $\mu = np$  fixed while  $n \rightarrow \infty$  this forces  $p \rightarrow 0$ . Then the limit for the pmf  $f(x)$  is given by

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

### 4.6.2 Poisson Distribution from Poisson Process

Consider an event that occurs in random points in time with the following conditions:

- **Independence** The number of occurrences of the event in non-overlapping intervals is independent
- **Individuality** For a small interval  $[T, T + \Delta t)$  we have

$$P[\text{Two or more events in } [T, T + \Delta t)] = o(\Delta t)$$

- **Homogeneity** events occur at a uniform rate  $\lambda$  over time.

Any process with these 3 conditions is called a **Poisson process**.

**4.10 Theorem.** Poisson Process In a poisson process with a rate of occurrence  $\lambda$  the number of occurrences  $X$  in a time interval  $t$  has a Poisson distribution. The pmf is given by

$$f(x) = P[X = x] = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

## 5 Expected Value and Variance

**5.1 Definition.** Expected Value Let  $X$  be a discrete random variable with range  $A$  and pmf  $f(x)$  then the expected value of  $X$  is

$$E(X) = \sum_{x \in A}^x f(x)$$

**5.2 Theorem.** *Law of unconscious statistician*

$$E(g(X)) = \sum_{x \in A} g(x)f(x)$$

**5.3 Theorem.** Linearity of Expectation

$$E(a \cdot g(X) + b) = aE(g(X)) + b$$

**5.4 Definition.** Variance Let  $\mu = E(X)$  then the variance is given by

$$\text{Var}(X) = E[(X - \mu)^2]$$

**5.5 Theorem.** *Variance properties* Let  $\sigma^2 = \text{Var}(X)$  then

$$(1) \quad \sigma^2 = E(X^2) - E(X)^2$$

$$(2) \quad \text{Var}(aX + b) = a^2 \text{Var}(X)$$

**5.6 Theorem.** Expected and Variance value for common distributions

$$(1) \quad X \sim \text{Bernoulli}(p) \text{ then } E(X) = p \text{ and } \text{Var}(X) = p(1 - p)$$

$$(2) \quad X \sim \text{Binomial}(n, p) \text{ then } E(X) = np \text{ and } \text{Var}(X) = np(1 - p)$$

$$(3) \quad X \sim \text{Geometric}(p) \text{ then } E(X) = \frac{1-p}{p} \text{ and } \text{Var}(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p}$$

$$(4) \quad X \sim \text{NB}(k, p) \text{ then } E(X) = \frac{k(1-p)}{p} \text{ and } \text{Var}(X) = \frac{k(1-p)}{p^2}$$

$$(5) \quad X \sim \text{Hypergeometric}(N, M, n) \text{ then } E(x) = \frac{nM}{N}$$

$$(6) \quad X \sim \text{Poisson}(\mu) \text{ then } E(x) = \mu \text{ and } \text{Var}(X) = \mu.$$

## 6 Multivariate Distributions

### 6.1 Joint and marginal pmf

Joint probability functions for more than 1 variables are defined as:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= P[X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_n = x_n] \\ &= P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] \end{aligned}$$

$f(x_1, \dots, x_n)$  is called the joint probability function of  $(X_1, \dots, X_n)$

**6.1 Definition.** Properties of Joint Probability function

$$(1) \quad \sum_{\text{all}(x_1, \dots, x_n)} f(x_1, \dots, x_n) = 1$$

$$(2) \quad f(x_1, \dots, x_n) \geq 0 \quad \forall (x_1, \dots, x_n)$$

**6.2 Definition.** Marginal probability function Marginal probability function for a single variable  $X$  or  $Y$  denoted by  $f_X(x)$  and  $f_Y(y)$  are

$$\begin{aligned} f_X(x) &= \sum_{\text{all}(y)} f(x, y) \\ f_Y(y) &= \sum_{\text{all}(x)} f(x, y) \end{aligned}$$

where  $f(x, y)$  is the joint probability function for  $X, Y$ .

#### 6.1.1 Conditional pmf

**6.3 Definition.** Conditional pmf The conditional *pmf* of  $X$  given  $Y = y$  is

$$f_X(x|y) = \frac{f(x, y)}{f_Y(y)}, \text{ Given } f_Y(y) > 0$$

#### 6.1.2 Independent random variables

**6.4 Definition.** Independent random variables Two random variables  $X, Y$  are independent if

$$P[X = x, Y = y] = P[X = x] \cdot P[Y = y]$$

Or  $f(x, y) = f_X(x)f_Y(y)$  for all  $(x, y)$ .

**6.5 Theorem.** Vandermonde's Convolution Formula

$$\binom{n+m}{k} = \sum_{j=0}^k \binom{n}{j} \binom{m}{k-j}$$

**7 Multivariate Distributions****7.1 Trinomial Distribution**

There are 3 possible outcomes  $A, B, C$  with

$$P[A] = p_1, P[B] = p_2, P[C] = p_3$$

$$p_1 + p_2 + p_3 = 1$$

The joint *pmf* when  $x_1 + x_2 + x_3 = n$  is given by

$$P[X_1 = x_1, X_2 = x_2, X_3 = x_3] = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

the marginal distribution is

$$X_1 \sim \text{Binomial}(n, p_1); X_2 \sim \text{Binomial}(n, p_2); X_3 \sim \text{Binomial}(n, p_3);$$

**7.2 Multinomial Distribution**

Each trial for  $k \geq 2$  has possible outcomes  $A_1, \dots, A_k$  with  $P[A_i] = p_i$ . Such that  $\sum p_i = 1$ . If we have

$$\sum X_i = n$$

Then the joint *pmf* is

$$(X_1, \dots, X_k) \sim \text{Multinomial}(n, p_1, \dots, p_k)$$

$$P[X_1 = x_1, \dots, X_k = x_k] = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

## 8 Multivariate Expectation

**8.1 Theorem.** The law of unconscious statistician If  $(X_1, X_2) \sim f(x_1, x_2)$  then

$$E(g(X_1, X_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) f(x_1, x_2)$$

**8.2 Theorem.** Properties of Expectation

- (1)  $E(X + Y) = E(X) + E(Y)$
- (2)  $E(aX + bY + c) = aE(X) + bE(Y) + c$
- (3) if  $X$  and  $Y$  are independent then

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)]$$

### 8.1 Covariance

**8.3 Definition.** Covariance

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E(XY) - (EX)(EY) \end{aligned}$$

**8.4 Definition.** Variance

$$\begin{aligned} \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ \text{Var}(aX + bY + c) &= a^2\text{Var}(X) + 2ab\text{Cov}(X, Y) + b^2\text{Var}(Y) \end{aligned}$$

**8.5 Theorem.** Multinomial co-variance If  $(X_1, \dots, X_k) \sim \text{Multinomial}(n, p_1, \dots, p_k)$  then

$$\text{Cov}(X_i, X_j) = -np_i p_j$$

### 8.2 Correlation coefficient

**8.6 Definition.**

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

$|\rho| \leq 1$  and  $|\rho| = 1$  if and only if  $Y = aX + b$ .

### 8.3 Expectation and Variance results

**8.7 Theorem.** General result

$$(1) \ E(\sum c_i X_i) = \sum c_i E(X_i)$$

$$(2) \ \text{Var}(\sum c_i X_i) = \sum c_i^2 \text{Var}(X_i) + \sum_{i < j} c_i c_j \text{Cov}(X_i, X_j)$$

$$(3) \ \text{If } X_1, \dots, X_n \text{ are independent then } \text{Var}(\sum c_i X_i) = \sum c_i^2 \text{Var}(X_i)$$

## 9 Continuous probability distributions

**9.1 Definition.** Probability density function (PDF)  $X$  has *pdf*, where

$$\begin{aligned} f(x) &\geq 0 & x &\in (-\infty, \infty) \\ P[a \leq X \leq b] &= \int_a^b f(x) dx & a &\leq b \end{aligned}$$

**9.2 Definition.** Cumulative distribution function (CDF)  $X$  has *cdf*, where

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(t) dt$$

$$(1) \quad 0 \leq F(x) \leq 1$$

$$(2) \quad F(x_1) \leq F(x_2) \text{ if } x_1 < x_2$$

$$(3) \quad \text{If } f(x) \text{ is continuous at } x \text{ then } \frac{dF(x)}{dx} = f(x)$$

### 9.1 Change of variables

Suppose  $X \sim f(x)$  and let  $Y = h(X)$ . If  $h(X)$  has a *unique* inverse or  $h(\cdot)$  is one-to-one then the *pdf* of  $Y$  is given by

$$g(y) = f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right|$$

*Proof.*

$$g(y) = \frac{d}{dy} F(h^{-1}(y)) = F'(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| = f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right|$$

□



## 9.2 Commonly used continuous distributions

### 9.3 Exponential distribution

$X \sim \text{EXP}(\lambda)$

(1) The *pdf* is  $f(x) = \frac{\lambda}{e^{\lambda x}}$   $x \geq 0, \lambda > 0$

(2) The *cdf* is

$$F(x) = \int_{-\infty}^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad x > 0$$

$$F(x) = 0 \quad x \leq 0$$

(3) Memory less property

$$P[X > t + s | X > s] = P[X > t]$$

(4) Can be used to model waiting time. Suppose  $\lambda$  is the intensity parameter of a Poisson process. Let  $X$  be the waiting time for the next event.

$$F(x) = 1 - P(X > x) = 1 - P[\text{No events in } [0, x]] = 1 - \frac{e^{-\lambda x} (\lambda x)^0}{0!} = 1 - e^{-\lambda x}$$

### 9.4 Gamma distribution

$X \sim \text{GAM}(\alpha, \beta)$

(1) The *pdf* is

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(\frac{-x}{\beta}\right)$$

### 9.5 Uniform distribution

$X \sim \text{UNIF}(a, b)$  then the *pdf* is

$$f(x) = \frac{1}{b - a}$$

the *cdf* is

$$F(x) = \frac{x - a}{b - a}$$

## 9.6 Beta distribution

Beta function:

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

(1) pdf is given by

$$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}$$

## 10 Expectation and Variance of continuous random variables

**10.1 Definition.** Expectation and variance

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$
$$\text{Var}(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx$$

## 11 Moments and moment generating function

**11.1 Definition.** If  $X$  is a random variable then the moment generating function is

$$M_X(t) = E(e^{tX}) \quad \text{Given it exists for } t \in (-h, h)$$

### 11.1 Joint MGF

**11.2 Definition.** Joint MGF For two random variables  $X, Y$  the joint MGF is

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E(e^{t_1 X + t_2 Y}) \\ &= \iint_{\mathbb{R}^2} e^{t_1 x} e^{t_2 y} f(x, y) dx dy \end{aligned}$$

if the joint expectation exists for  $t_1 \in (-h_1, h_1)$  and  $t_2 \in (-h_2, h_2)$   $M_X(t) = M_{X,Y}(t, 0)$  and  $M_Y(t) = M_{X,Y}(0, t)$

**11.3 Theorem.** MGF of linear function Suppose  $X$  has mgf  $M_X(t)$  for  $t \in (-h, h)$ . Let  $Y = aX + b$  then

$$M_Y(t) = e^{bt} M_X(at) \quad t \in \left( -\frac{h}{|a|}, \frac{h}{|a|} \right)$$

## 12 Properties of MGF

**12.1 Theorem.** Moments Suppose  $X$  has MGF  $M_X(t)$ . Then  $M_X(0) = 1$  and

$$E(X^k) = M_X^{(k)}(0)$$

**12.2 Theorem.** Joint Moments Suppose  $X, Y$  are random variables with joint MGF  $M_{X,Y}(t_1, t_2)$ . Then

$$E(X^j Y^k) = \left. \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M_{X,Y}(t_1, t_2) \right|_{(t_1, t_2) = (0, 0)}$$

**12.3 Theorem.** MGF Independence Two random variables  $X, Y$  are independent **if and only if**

$$M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$$

**12.4 Theorem.** Sum of independent random variables If  $Y = X_1 + X_2$  then

$$M_Y(t) = M_1(t)M_2(t)$$

## 13 Convergence in probability

**13.1 Definition.** Convergence in probability  $X_n \xrightarrow{p} b$  if for any  $\varepsilon > 0$  we have

$$\lim_{n \rightarrow \infty} P[|X_n - b| \geq \varepsilon] = 0$$

**13.2 Theorem.** Convergence of Sum and product If  $X_n \xrightarrow{p} a$  and  $Y_n \xrightarrow{p} b$  then

$$X_n + Y_n \xrightarrow{p} a + b$$

$$X_n Y_n \xrightarrow{p} ab$$

**13.3 Theorem.** Markov's inequality For random variable  $X$  with finite  $E(|X|^k)$ ,

$$P[|X| \geq c] \leq \frac{E[|X|^k]}{c^k}$$

### 13.1 Weak Law of large numbers

**13.4 Theorem.** Weak Law of large numbers Suppose  $X_i$ 's are **iid** with  $EX_i = \mu$  and  $\text{Var}X_i = \sigma^2 < \infty$  then

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

$$\bar{X} \xrightarrow{p} \mu$$

## 14 Convergence in distribution

**14.1 Definition.** Convergence in distribution  $X_n \xrightarrow{d} X$  if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{All all continuity points of } F$$

where  $F_n(x) = P[X_n \leq x]$  and  $F$  is the *CDF* of  $X$

### 14.1 e limit

**14.2 Theorem.** *e* limit If  $b$  is a real constant and  $\lim_{n \rightarrow \infty} \psi(n) = 0$  then

$$\lim_{n \rightarrow \infty} \left[ 1 + \frac{b}{n} + \frac{\psi(n)}{n} \right]^n = e^b$$

### 14.2 MGF for limiting distributions

**14.3 Theorem.** MGF Convergence theorem Let  $X_1, X_2, \dots, X_n$  and let  $M_1(t), M_2(t), \dots, M_n(t), \dots$  be the MGF. Let  $X$  be a random variable with MGF  $M(t)$  if there exists  $h > 0$  such that

$$\lim_{n \rightarrow \infty} M_n(t) = M(t) \quad t \in (-h, h)$$

then  $X_n \xrightarrow{d} X$

## 15 Central limit theorem

**15.1 Theorem.** Central limit theorem Suppose  $X_i$  are *iid* with  $EX_i = \mu$  and  $\text{Var}X_i = \sigma^2 < \infty$  then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1)$$

Equivalently

$$(1) \ S_n = \sum X_i \xrightarrow{d} N(n\mu, n\sigma^2)$$

$$(2) \ \bar{X} = \frac{1}{n} \sum X_i \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$$

## CDF Not in notes

Name	$PDF$	$E(X)$	$\text{Var}(X)$	MGF ( $M(t)$ )
Exponential ( $X \sim EXP(\lambda)$ )	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1-\frac{t}{\lambda}}$ for $t < \lambda$
Gamma ( $X \sim Gamma(\alpha, \beta)$ )	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\alpha \cdot \beta$	$\alpha \cdot \beta^2$	$\frac{1}{(1-\beta t)^\alpha}$ for $t < \frac{1}{\beta}$
Beta ( $X \sim Beta(\alpha, \beta)$ )	$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	-



## Useful

IID for distribution

# 1 If  $Y \sim BN(n, p)$  then

$$Y = \sum_{i=1}^n X_i \qquad X_i \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$$

# 2 If  $Y \sim \text{POI}(n)$  then

$$Y = \sum_{i=1}^n X_i \qquad X_i \stackrel{\text{iid}}{\sim} \text{POI}(1)$$

# 3 If  $Y \sim \text{GAMMA}(\alpha, \beta)$  then

$$Y = \sum_{i=1}^n X_i \qquad X_i \stackrel{\text{iid}}{\sim} \text{GAMMA}(1, \beta)$$

# 4 If  $Y \sim NB(k, p)$  then

$$Y = \sum_{i=1}^k X_i \qquad X_i \stackrel{\text{iid}}{\sim} \text{Geo}(p)$$

Assignment Theorems:

**15.2 Theorem. A4Q7** If  $\{X_1, X_2, \dots, X_n, \dots\}$  is a sequence of random variables and  $E(X_n) = \mu$  and  $\text{Var}(x_n) = \sigma_n^2$  with

$$\lim_{n \rightarrow \infty} \mu_n = \mu$$

$$\lim_{n \rightarrow \infty} \sigma_n^2 = 0$$

Then

$$X_n \xrightarrow{p} X = \mu$$