Wasserman (2005) 23/04/2020

Ch3 - Expectation

3.1 Expectation of a random variable

- mean/expectation of an r.v. X is the average value of X

3.1 Definition

- The expected value/mean/1st moment of X is defined as:-

$$\mathbb{E}[X] = \int x \, dF(x) = \begin{cases} \sum_x x f(x) & X \text{ discrete} \\ \int x f(x) \, dx & X \text{ continuous} \end{cases} \quad (3.1)$$

- This assumes that the integral is well-defined
- We use following notation:-

$$\mathbb{E}[X] = \mathbb{E}X = \int x \, dF(x) = \mu = \mu_X \quad (3.2)$$

- $\mathbb{E}(X)$ is a one-number summary of distri
- Heuristically (at this stage); $\mathbb{E}[X]$ is average $\sum_{i=1}^{n} X_i / n$ of a large no. of IID draws $X_1, \dots, X_n$
- $\mathbb{E}(X) \approx \sum_{i=1}^{n} X_i / n$ is more than a heuristic (LLN Ch 5)
- $\int x \, dF(x)$ is convenient notation (in this course); but has a precise meaning in real analysis courses.

Q: well defined $\mathbb{E}[X]$: $\mathbb{E}[X]$ exists if $\int_x |x| \, dF_X(x) < \infty$ ($\mathbb{E}[|X|] < \infty$); otherwise expectation does not exist

3.5 Example

Q: As $\mathbb{E}[|X|] \not< \infty$ for Cauchy distri, mean does not exist

- Simulation of Cauchy distri many times + average; average does not settle down
- Thick tails, common extreme observations.

- Implicitly assume for this course that expectations exist
- For transformations e.g. $Y = r(X)$; computing $\mathbb{E}[Y]$?

1. Find $f_Y(y)$, compute $\int y f_Y(y) \, dy$

OR

3.6 Theorem (lazy statistician)

- Let $Y = r(X)$; then
$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \int r(x)\, dF_X(x) = \int r(x) f_X(x)\, dx \qquad (3.3)$$

- makes intuitive sense; e.g. payouts on outcome of gambling!
- e.g. Play a game, draw X, payout $r(X)$
- Average income $r(x)$ times chance that $X=x$; summed / integrated over all x.
- A special case: (indicators, exp)
- Let A be an event and let $r(x) = I_A(x)$ where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$ (2)
- $\mathbb{E}[I_A(X)] = \int I_A(x) f_X(x)\, dx = \int_A f_X(x)\, dx = P(X \in A)$
- probability is a special case of expectation

Ex 3.7

(!): Note that when computing definite integrals for $\mathbb{E}[r(X)]$; the relevant bounds are the domain of $r(X)$; or the ones governing PDF.

Ex 3.8

(!) note the piecewise definition of the transformation $r(X)$

(!): very interesting use of breaking/constructing problem probab.

- Transformations of several r.vs.
- If $Z = r(X,Y)$ then
$$\mathbb{E}[Z] = \mathbb{E}[r(X,Y)] = \iint r(x,y)\, dF(x,y) = \iint r(x,y) f_{X,Y}(x,y)\, dx\, dy \qquad (3.4)$$

3.9 Example

- Let (X,Y) have a jointly uniform distn on unit square i.e.
$$f_{X,Y}(x,y) = \begin{cases} 1 & 0<x<1,\ 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

- Let $Z = r(X,Y) = X^2 + Y^2$, then
$$\mathbb{E}[Z] = \iint r(x,y)\, dF(x,y) = \int_0^1 \int_0^1 (x^2+y^2)\, dx\, dy = \int_0^1 x^2\, dx + \int_0^1 y^2\, dy$$

$= [\frac{1}{3}x^3]_0^1 + [\frac{1}{3}y^3]_0^1 = \frac{2}{3}$

- Note that this is the same as :-

$$\int_0^1\int_0^1 x^2+y^2\,dx\,dy = \int_0^1 [\tfrac{1}{3}x^3 + xy^2]_0^1\,dy = \int_0^1 \tfrac{1}{3}+y^2\,dy = [\tfrac{1}{3}y + \tfrac{1}{3}y^3]_0^1 = \tfrac{2}{3}$$

- The k^{th} moment of X is defined as $\mathbb{E}[X^k]$; and assumes that $\mathbb{E}[|X|^k] < \infty$

3.10 Theorem

- If the k^{th} moment exists (i.e. $\mathbb{E}[|X|^k] < \infty$), and if $j < k$, then the j^{th} moment exists

Proof ? - (W) (*): need to review; don't fully understand - supplementary pages

- The k^{th} central moment is defined as $\mathbb{E}[(X-\mu)^k]$

3.2 Properties of Expectations

3.11 Theorem (Linearity, in stats) not necess indep. (*)

- If $X_1, \dots, X_n$ are r.v.s. and $a_1, \dots, a_n$ are constants then:-

$$\mathbb{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathbb{E}[X_i] \quad (3.5)$$

- 3.12 Example
- $X \sim Binomial(n,p)$
- Mean of X?
- Appeal to definition: $\mathbb{E}[X] = \int x\, dF_X(x) = \sum_x x f_X(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^x$
- Not easy to evaluate.
- Instead, note $X = \sum_{i=1}^{n} X_i$ where $X_i = 1$ if i^{th} toss heads; $X_i = 0$ if i^{th} toss tails
- Then $\mathbb{E}(X_i)$ (individual expectation) $= (p \times 1) + ((1-p) \times 0) = p$
- $\mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}[X_i] = np$

(*): Decompose X into a series of Bernoulli r.v.s. X_i

3.13 Theorem (Expectation op distributes over product for independent r.v.s)

- Let $X_1, \dots, X_n$ be independent random variables. Then,

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i) \quad (3.6)$$

(6n): Summation rule does not require independence; but multiplication does

3.3. Variance and covariance

- variance measures 'spread of distr'
- we do not normally use $\mathbb{E}(X-\mu)$ as $\mathbb{E}(X-\mu) = \mu-\mu=0$
- sometimes $\mathbb{E}[|X-\mu|]$, but more commonly, variance

3.14 Definition (variance, sd)

- Let X be an r.v. with mean μ.
- The variance of X, denoted σ^2 or σ_X^2 or $\mathbb{V}(X)$ or $\mathbb{V}X$:-

$$\sigma^2 = \mathbb{E}[(X-\mu)^2] = \int (x-\mu)^2 \, dF_X(x) \qquad (3.7)$$

- Assuming this expectation exists
- The standard deviation $sd(X) = \sqrt{\mathrm{Var}(X)}$ is also denoted σ and σ_X.

3.15 Theorem (variance prop.)

- Assuming the variance is well-defined (?)
- It has following properties:-
  1. $\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2$
  2. If a and b are constants; $\mathbb{V}(aX+b) = a^2\,\mathbb{V}(X)$
  3. If $X_1,\dots,X_n$ are independent and $a_1,\dots,a_n$ constants:-

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2\, \mathrm{Var}(X_i) \qquad (3.8)$$

3.16 Example

(W): Review intuition precisely (underlined)

- If $X_1,\dots,X_n$ are r.v.s, we define the sample mean :-

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (3.9)$$

- And sample variance:-

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \qquad (3.10)$$

- These are sampling estimators of their population analogues

· If $Y = aX + b$ for some constants a, b, then $\rho(X,Y) = 1$ if $a > 0$ and $\rho(X,Y) = -1$ if $a < 0$ (i.e. Y and X are linearly related with former being a linear transformation of latter)

· If X and Y are independent then $cov(X,Y) = \rho = 0$;

· NB: However, in general, the converse is not true

3.20 Theorem

· $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$

· $var(X-Y) = var(X) + var(Y) - 2cov(X,Y)$

· more generally, for r.v.s. $X_1, \dots, X_n$:

$$var\left(\sum_i a_i X_i\right) = \sum_i a_i^2 var(X_i) + 2\sum\sum_{i<j} a_i a_j cov(X_i, X_j)$$

3.4 Expectation, Variance of Important Random Variables

- Important mean and variances ⟶ see Wasserman.
- notable:
- Multivariate models:
- Random vector $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_k] \end{pmatrix}$

- variance covariance matrix:

$$\underline{\Sigma} = var(\underline{X}) = \begin{bmatrix} var(X_1) & cov(X_1,X_2) & \cdots & cov(X_1,X_k) \\ cov(X_2,X_1) & var(X_2) & \cdots & cov(X_2,X_k) \\ & \vdots & \ddots & \\ cov(X_k,X_1) & cov(X_k,X_2) & & var(X_k) \end{bmatrix}$$

i.e. $\Sigma_{ij} = cov(X_i, X_j)$

$\Sigma_{ii} = cov(X_i, X_i) = var(X_i)$

· $X \sim Multinomial(n, \underline{p})$ then

· $\mathbb{E}[X] = n\underline{p} = n(p_1, \dots, p_k)$ $var(\underline{X}) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & & \ddots & \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{pmatrix}$

3.17 Theorem (ST)

- let $X_1, \dots, X_n$ be iid and let $\mu = \mathbb{E}(X_i)$ $\sigma^2 = Var(X_i)$; then:-

$$\mathbb{E}[\bar{X}_n] = \mu \quad \mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n} \quad \mathbb{E}[S_n^2] = \sigma^2$$

- make sure you understand this well; it's a little meta.
- $\bar{X}_n$ and S_n^2 are statistics / estimators which are functions of the data; and hence are random, because the data is random (in freq paradigm); even though they are deterministic functions (ST).
- The above relates the 1st, 2nd moments of the sample mean estimator, and 1st moment of sample variance estimator; i.e. properties of the sampling distribution (a distribution on the r.v. $\bar{X}_n, S_n^2$) with that of the underlying IID r.v.s on which those estimators are constructed.
- see lecture notes also; really important distinctions

- If X and Y are random variables; then the covariance and correlation between X and Y measure how strong the linear rel. is between X and Y.

3.18 Definition (covariance, correlation)

- let X and Y be random variables with means μ_X and μ_Y and sds. σ_X and σ_Y
- define the covariance between X and Y:-

$$cov(X,Y) = \mathbb{E}\left((X-\mu_X)(Y-\mu_Y)\right) \qquad (3.11)$$

- And the correlation by:-

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad (3.12)$$

(in effect σ_X, σ_Y product as like 'normalis')

3.19 Theorem

- The covariance satisfies:-

$$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

- The correlation satisfies

$$-1 \leq \rho(X,Y) \leq 1$$

- Understanding $var(\underline{X})$ for $\underline{X} \sim Multinomial(n, \underline{p})$
- Any one component of $\underline{X}$, $X_i \sim Binomial(n, p_i)$
- $\therefore \mathbb{E}[X_i] = np_i \quad var[X_i] = np_i(1-p_i)$
- Note that $X_i + X_j \sim Binomial(n, p_i + p_j)$?
- Thus $var(X_i + X_j) = n(p_i + p_j)(1 - [p_i + p_j])$ (I)
- Also, $var(X_i + X_j) = var(X_i) + var(X_j) + 2cov(X_i, X_j) = np_i(1-p_i) + np_j(1-p_j) + 2cov(X_i, X_j)$ (II)
- (I) = (II), solve $\Rightarrow cov(X_i, X_j) = -np_ip_j$
- Lemma for finding means and variances of linear combinations of multivariate random vectors

3.21 Lemma

- If $\underline{a}$ is a vector and $\underline{X}$ is a random vector with mean $\underline{\mu}$ and variance $\underline{\Sigma}$, then $\mathbb{E}[\underline{a}^T \underline{X}] = \underline{a}^T \underline{\mu}$ and $var(\underline{a}^T \underline{X}) = \underline{a}^T \underline{\Sigma}\, \underline{a}$
- If A is a matrix then:-

$$\mathbb{E}[\underline{A}\underline{X}] = \underline{A}\underline{\mu} \text{ and } var(\underline{A}\underline{X}] = \underline{A}\underline{\Sigma}\underline{A}^T$$

3.5 Conditional Expectation

- Suppose X and Y are r.v.s.
- What is the mean of X among those times when $Y = y$?
- Compute mean of X as before but subst. $f_{X|Y}(x|y)$ in place of $f_X(x)$.

3.24 Definition

- The conditional expectation of X given $Y = y$ is:-

$$\mathbb{E}[X|Y=y] = \begin{cases} \sum x f_{X|Y}(x|y) \cancel{dx} & \text{discrete} \\ \int x f_{X|Y}(x|y)dx & \text{continuous} \end{cases} \quad (3.13)$$

?

- If $r(x,y)$ is a function of x and y then:-

$$\mathbb{E}[r(X,Y)|Y=y] = \begin{cases} \sum r(x,y) f_{X|Y}(x|y) \cancel{dx} & \text{discrete} \\ \int r(x,y) f_{X|Y}(x|y)dx & \text{continuous} \end{cases} \quad (3.14)$$

(OT): $\mathbb{E}[X]$ is just a number (output-side)

(W) $\mathbb{E}[X|Y=y]$ is a function of y (that which we are conditioning on)

· before we observe Y, we do not know value of $\mathbb{E}[X|Y=y]$; so it is a random variable $\mathbb{E}[X|Y]$

· Alternatively $\mathbb{E}[X|Y]$ is the random variable whose value is $\mathbb{E}[X|Y=y]$ when $Y=y$.

· Similarly; $\mathbb{E}[r(X,Y)|Y]$ is the r.v. whose value is $\mathbb{E}[r(X,Y)|Y=y]$ when $Y=y$.

· very confusing point (I agree); always get hazy with this... no longer.

· example 3.23.

- suppose we draw $X \sim \text{Unif}(0,1)$

- After we observe $X=x$, we draw $Y|X=x \sim \text{Unif}(x,1)$ → (W): Recall earlier confus. about this notion.

- Intuitively, expect $\mathbb{E}[Y|X=x] = \frac{1+x}{2}$ (do we?)

- $f_{Y|X}(y|x) = \frac{1}{1-x}$ for $x<y<1$ → $f_X(x) = \begin{cases} 1 & 0<x<1 \\ 0 & \text{otherwise} \end{cases}$ $f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & x<y<1 \\ 0 & \text{oth.} \end{cases}$

- $\mathbb{E}[Y|X=x] = \int_x^1 y f_{Y|X}(y|x)\,dy = \frac{1}{1-x}\int_x^1 y\,dy = \frac{1}{1-x}\left[\frac{1}{2}y^2\right]_x^1$

$$= \frac{1}{1-x}\left(\frac{1}{2}-\frac{1}{2}x^2\right) = \frac{1-x^2}{2(1-x)} = \frac{(1+x)(1-x)}{2(1-x)} = \frac{1+x}{2}$$

- Thus $\mathbb{E}[Y|X] = \frac{1+X}{2}$

- This an r.v. $\frac{1+X}{2}$, with value/realisation $\mathbb{E}[Y|X=x] = \frac{1+x}{2}$ once $X=x$ has been observed

(OT): The distinction between $\mathbb{E}[X|Y]$ and $\mathbb{E}[X|Y=y]$ is based on whether or not what is being conditioned has already been observed

(W): check you fully internalised/assimilated this.

3.24 Theorem (Iterated expectations)

For r.v.s X and Y, assuming expectations exist; we have that

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] \quad \text{and} \quad \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

More generally, for any function $r(x,y)$, we have:-

$$\mathbb{E}[\mathbb{E}[r(X,Y)|X]] = \mathbb{E}[r(X,Y)]$$

(w): I don't like this style notation; prefer $\mathbb{E}_{x\sim f_X(x)}$ as it is clear what the underlying probability distribution over which expectations are taken is.

Proof of 1st eq.:

- via conditional exp. defn; $f(x,y) = f(x)f(y|x)$;

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int \mathbb{E}[Y|X=x]\, f_X(x)\,dx = \iint y\, f_{Y|X}(y|x)\,dy\, f_X(x)\,dx$$

$$= \iint y\, f(y|x) f(x)\,dx\,dy = \iint y\, f(x,y)\,dx\,dy = \mathbb{E}_{X,Y\sim f_{X,Y}(x,y)}[Y]$$

- (w): see it is not entirely clear whether $\mathbb{E}[Y]$ is $\mathbb{E}_{Y\sim f_Y(y)}[Y]$ or $\mathbb{E}_{X,Y\sim f_{X,Y}(x,y)}[Y]$

- (w): needs further clarifying - there are some questions here I need to articulate

3.25 Example

- Consider 3.23.
- How can we compute $\mathbb{E}_{X,Y\sim f_{X,Y}(x,y)}[Y]$?
- Method 1: Find joint density $f_{X,Y}(x,y)$, compute $\mathbb{E}[Y] = \iint y\, f(x,y)\,dx\,dy$
- Method 2: $\mathbb{E}[Y|X] = \frac{1+X}{2}$;

$$\mathbb{E}[Y] = \mathbb{E}_{X\sim f_X(x)}\left[\mathbb{E}_{Y|X\sim f_{Y|X}(y|x)}[Y|X]\right] = \mathbb{E}_{X,Y\sim f_{X,Y}(x,y)}\left[\frac{1+X}{2}\right]$$

$$= \frac{1+\mathbb{E}_{X,Y\sim f(x,y)}[X]}{2} = \frac{1+(1/2)}{2} = \frac{3}{4}$$

3.26 Definition

- The conditional variance is defined as:-

$$\mathrm{var}(Y|X=x) = \int (y-\mu(x))^2 f(y|x)\,dy \qquad (3.17)$$

where $\mu(x) = \mathbb{E}[Y|X=x]$

3.2.7. Theorem

- For random variables X and Y,

$$\text{var}(Y) = \mathbb{E}[\text{var}(Y|X)] + \text{var}[\mathbb{E}[Y|X]]$$

3.28 Example

- Draw a county at random from U.S.
- Draw n people at random from that county
- Define X - no. of those people with a disease

Q - proportion of people in that county with the disease (varies from county to county)

- Given $Q=q$; we have that $X \sim \text{Binomial}(n, q)$

Ex: Ammo to undecided Flaxman

$\Rightarrow \mathbb{E}[X|Q=q] = nq \quad \text{var}(X|Q=q) = nq(1-q)$

- Suppose that Q has a $\text{Unif}(0,1)$ distn
- This is a distribution that is constructed in stages $\longrightarrow$ hierarchical model

(A11)

$Q \sim \text{Uniform}(0,1)$

$X|Q=q \sim \text{Binomial}(n, q)$

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Q]] = \mathbb{E}[nQ] = n\mathbb{E}(Q) = n/2$

$\longrightarrow$ note point earlier; Q not yet obs.

- Compute variance of X:

$$\text{var}(X) = \underbrace{\mathbb{E}[\text{var}(X|Q)]}_{(i)} + \underbrace{\text{var}[\mathbb{E}[X|Q]]}_{(ii)}$$

(i): $\mathbb{E}[\text{var}[X|Q]] = \mathbb{E}[nQ(1-Q)] = n\mathbb{E}[Q(1-Q)] = n\int_0^1 q(1-q)\,dq = n\left[\frac{1}{2}q^2 - \frac{1}{3}q^3\right]_0^1$

$= \frac{n}{6}$

(ii) $\text{var}[\mathbb{E}[X|Q]] = \text{var}[nQ] = n^2\text{var}[Q] = n^2\int_0^1 (q-\frac{1}{2})^2\,dq = n^2\int_0^1 q^2 - q + \frac{1}{4}\,dq$

$= n^2\left[\frac{1}{3}q^3 - \frac{1}{2}q^2 + \frac{1}{4}q\right]_0^1 = n^2\left(\frac{1}{3} - \frac{1}{2} + \frac{1}{4}\right) = \frac{n^2}{12}$

3.6 Moment Generating Functions

- For finding moments, distn of sums of r.v.s, proofs.

3.29 Definition

- The moment generating function, MGF, or laplace transform of X is defined by:-

$$\psi_X(t) = \mathbb{E}[e^{tX}] = \int e^{tx}\, dF(x)$$

- where t varies over the real nos $\mathbb{R}$.

- Assume MGF is well-defined for all t in some open interval around $t=0$
- (i): When the MGF is well-defined we interchange operations of 'differentiation' and 'expectation': i.e. some kind of similarity of operators:-

$$\psi'(0) = \left[\frac{d}{dt}\mathbb{E}[e^{tX}]\right]_{t=0} = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right]_{t=0} = \mathbb{E}\left[Xe^{tX}\right]_{t=0} = \mathbb{E}[X]$$

(ii): By taking k derivatives, we conclude that

$\psi^{(k)}(0) = \mathbb{E}[X^k] \longrightarrow$ a method for computing moments of a distribution

3.30 example — i.e. $\lambda(?)$ $f_X(x) = e^{-x}$ $x > 0$

- Let $X \sim Exp(1)$. For any $t < 1$

$$\psi_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx}e^{-x}\,dx = \int_0^\infty e^{(t-1)x}\,dx = \left[\frac{1}{(t-1)}e^{(t-1)x}\right]_0^{x\to\infty}$$

(iv): limit behaviour review $= \frac{1}{1-t}$

- Divergent integral if $t \geq 1$
- $\psi_X(t) = \frac{1}{1-t}$ $\forall t < 1$
- $\psi_X'(0) = -(1-t)^{-2} \cdot -1 = \frac{1}{(1-t)^2}\Big|_{t=0} = 1 = \mathbb{E}[X]$
- $\psi''_X(0) = -2(1-t)^{-3} \cdot -1 = \frac{2}{(1-t)^3}\Big|_{t=0} = 2$
- $Var(X) = \mathbb{E}(X^2) - \mu^2 = \psi''_X(0) - (\psi'_X(0))^2 = 1$

3.31 Lemma (MGF properties) — where ψ_i is the MGF of X

1. If $Y = aX + b$; $\psi_Y(t) = e^{bt}\psi_X(at)$
2. If $X_1, \dots, X_n$ are independent and $Y = \sum_i X_i$, then $\psi_Y(t) = \prod_i \psi_i(t)$

3.32. Example

- let $X \sim Binomial(n,p)$ ✓ IID
- we know that $X = \sum_{i=1}^{n} X_i$ where $P(X_i = 1) = p$; $P(X_i = 0) = 1-p$ (ie. sum of Bernoullis)
- $\psi_i(t)$ (MGF of each Bernoulli X_i) $= \mathbb{E}[e^{tX_i}] = pe^t + (1-p) = pe^t + q$ where $q = (1-p)$
- $\psi_X(t) = \prod_i \psi_i(t) = (pe^t + q)^n$

· Recall, X and Y are equal in distribution if they have the same distri fn (CDF) and we write $X \stackrel{d}{=} Y$.

3.33 Theorem

Q: Review

- let X and Y be r.v.s.
- If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then $X \stackrel{d}{=} Y$

3.34 Example

- let $X_1 \sim Binomial(n_1, p)$, $X_2 \sim Binomial(n_2, p)$ and independent
- let $Y = X_1 + X_2$, then: (via L.3.3.1.)

$\psi_Y(t) = \psi_1(t)\psi_2(t) = (pe^t+q)^{n_1}(pe^t+q)^{n_2} = (pe^t+q)^{n_1+n_2}$

- And latter is MGF of $Binomial(n_1+n_2, p)$ distri.
- As MGF characterises the distribution (no other r.v. with same MGF)
- we conclude $Y \sim Binomial(n_1+n_2, p)$
- Moment Generating Functions for common distri.

Distri	MGF $\psi(t)$
Bernoulli(p)	$pe^t + (1-p)$
Binomial(n,p)	$(pe^t + (1-p))^n$
Poisson(λ)	$e^{\lambda(e^t - 1)}$
Normal(μ, σ^2)	$\exp\{\mu t + \frac{\sigma^2 t^2}{2}\}$
Gamma(α, β)	$\left(\frac{1}{1-\beta t}\right)^\alpha$ for $t < 1/\beta$

3.35 Example

- Let $Y_1 \sim \text{Poisson}(\lambda_1)$, $Y_2 \sim \text{Poisson}(\lambda_2)$ and independent.
- Moment gen fn of $Y = Y_1 + Y_2$ is $\psi_Y(t) = \psi_{Y_1}(t)\psi_{Y_2}(t)$

$$\psi_Y(t) = e^{\lambda_1 e^t - 1} e^{\lambda_2 e^t - 1} = e^{(\lambda_1+\lambda_2)e^t - 1}$$

- MGF for $\text{Poisson}(\lambda_1 + \lambda_2)$
- Proved that sum of 2 ~~poissons~~ independent Poisson distn r.v.s. ~~Ba~~ has Poisson distn.

3.7 Appendix

- Expectations as integral
- The integral of a measurable function $r(x)$ is defined thus:-
- Suppose r is 'simple' i.e. takes finitely many values $a_1, \ldots, a_k$ over a partition $A_1, \ldots, A_k$.
- Define: $$\int r(x)\,dF(x) = \sum_{i=1}^{k} a_i P(r(X) \in A_i)$$
- The integral of a positive measurable function r is defined by:-

$\int r(x)\,dF(x) = \lim_i \int r_i\,dF(x)$ where r_i is a sequence of simple functions such that $r_i(x) \le r(x)$ and $r_i(x) \to r(x)$ as $i \to \infty$

- Not depend on particular sequence
- Integral of measurable function r is defined to be:- $\int r(x)\,dF(x)$

$= \int r^+(x)\,dF(x) - \int r^-(x)\,dF(x)$, assuming both integrals are finite

- where $r^+(x) = \max\{r(x), 0\}$ and $r^-(x) = -\min\{r(x), 0\}$