week 6

RECALL

2.2: Expectation

Def: The expectation of a RV $g(X)$ is

$$E[g(X)] := \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\,dx, & X\text{-continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x), & X\text{-discrete} \end{cases} \quad (*)$$

Remarks: • Also known as 'expected values', 'average' of a RV or of the prob. dist. of the RV $g(X)$.

Ex (Expectation): what is the expectation of X?
If $X \in \mathbb{R}$, then by letting $g(X) := x$

$$E(X) = \begin{cases} \int x f(x)\,dx, & X\text{-cont.} \\ \sum_{x \in \mathcal{X}} x \cdot f_X(x), & X\text{-discrete} \quad (**) \end{cases}$$

Note: If the domain of $\mathcal{X}$ is not a subset of $\mathbb{R}$ (Euclidean space), then (**) may be invalid but a notation of $E(X)$ may still be designed via the expectations of $\{Eg(X)\}$ in (*)

Note: Expectation is associated with a distribution

Ex (Cauchy Expectation Nuances): Let $X \sim \text{Cauchy}$; $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, $x \in \mathcal{X} = \mathbb{R}$.

check $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{1}{1+x^2}\,dx = 1$;

letting $x = \tan(\theta)$, $\frac{dx}{d\theta} = \frac{d}{d\theta}(\tan(\theta))$ (var-sub) $\Rightarrow dx = \sec^2(\theta)\,d\theta$] $\frac{d}{d\theta}\tan\theta = \sec^2\theta$ *FS

For $x = \infty \Rightarrow \theta = \tan^{-1}(\infty) = \pi/2$
$x = -\infty \Rightarrow \theta = \tan^{-1}(\infty) = -\pi/2$

$$= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1}{1+\tan^2\theta}\sec^2\theta\,d\theta, \quad \frac{1}{1+\tan^2\theta} = \sec^2\theta \,] \text{ *FS}$$

$$= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} d\theta = \frac{1}{\pi}\left[\frac{\pi}{2} + \frac{\pi}{2}\right] = 1.$$

$$E|X| = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{|x|}{1+x^2}\,dx = \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2}\,dx \quad \text{by symmetric cauchy RV}$$

For any $M > 0$:

$$\lim_{m\to\infty} \frac{2}{\pi}\int_0^m \frac{x}{1+x^2}\,dx$$

let $u = 1+x^2$, $du = 2x\,dx \Rightarrow \frac{1}{2}du = x\,dx$ } u-sub

$$= \lim_{m\to\infty} \frac{2}{\pi}\frac{1}{2}\int_0^m \frac{1}{u}\,du = \lim_{m\to\infty} \frac{1}{\pi}\left[\log(1+x^2)\right]\Big|_0^m$$

$$= \lim_{m\to\infty} \frac{1}{\pi}\log(1+m^2) = \infty$$

$$E(X) := \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{0} x f_X(x)\,dx + \int_0^{\infty} x f_X(x)\,dx$$

$$= \lim_{m\to\infty}\int_{-m}^{0} \cdot + \lim_{n\to\infty}\int_0^n \cdot = \lim_{\substack{m\to+\infty \\ n\to+\infty}} \frac{-1}{2\pi}\log(1+m^2) + \frac{1}{2\pi}\log(1+n^2) \text{ DNE (undefined)}$$

Remarks: • a proper expectation exists only when the positive and negative part of integral are finite $\left(\int |x| f_X(x)\,dx < \infty\right)$
• $E(X)$ is undefined since $\int x f_X(x)\,dx$ is not absolutely convergent

Prop.

Let X-RV with dist. P_X.

Let $X_1, \dots, X_n$ be n mutually independent RV's that have identical distributions as X, i.e. $X_1, \dots, X_n$ is an n-iid sample of P_X

Def (Empirical Distribution): of $\{X_1, \dots, X_n\}$ is a prob. dist., denoted by P_n, s.t.
if $Y \sim P_n$ then $y \in \mathcal{Y} = \{X_1, \dots, X_n\}$ and
$P(Y = X_i) = \frac{1}{n}$ for $i = 1, \dots, n$

Remarks: Thus, Y is discrete (regardless of X) and

$$E(Y) = \sum_{y \in \mathcal{Y}} P(Y=y) = y = \frac{1}{n}(X_1 + \dots + X_n) \neq E(X) = \int x f_X(x)\,dx$$

• Colloquially, this $E(Y)$ is also called the "average" of the (data) sample $X_1, \dots, X_n$.

Linearity of Expectation

Thm. Let X, Y : real-valued RV's for which the expectations exist. (~)

$\boxed{E(aX+bY+c) = aE(X)+bE(Y)+c}$ for $a,b,c \in \mathbb{R}$

Proof (~). Assume X and Y may be expressed as functions of a RV Z, i.e.

$$\begin{cases} X = g_1(Z) \\ Y = g_2(Z) \end{cases} \quad (*)$$

Suppose Z is a continuous RV w/ f_Z. Then

$E(aX+bY+c)$

$= E[ag_1(Z) + bg_2(Z) + c]$

$= \int (ag_1(z) + bg_2(z) + c) f_Z(z)\, dz$

$= a\int g_1(z) f_Z(z)\, dz + b\int g_2(z) f_Z(z)\, dz + c\int f_Z(z)\, dz$

$= aEg_1(Z) + bEg_2(Z) + c$

$= aE(X) + bE(Y) + c.$

The proof proceeds similarly if Z is discrete. □ (incomplete since $*$)

EX (Binomial): $X \sim \text{Binomial}(n,p)$

Let $Y_i = \begin{cases} 1, & \text{coin turns H at the } i^{th} \text{ toss} \\ 0, & \text{otherwise} \end{cases}$

Then $E(Y_i) = 1\cdot p + 0\cdot(1-p) = p$.

Here, $X = Y_1 + \cdots + Y_n$

and by linearity of expectation

$E(X) = E(Y_1) + \cdots + E(Y_n) = np.$

Corollary: ii) If $X \geq 0$ almost surely, i.e. $P(X\geq 0)=1$ then $E(X) \geq 0$

iii) If $X \geq Y$ almost surely, i.e. $P(X \geq Y) = 1$ then $E(X) \geq E(Y)$

iv) If $P(X \in [a,b]) = 1$ then $E(X) \in [a,b]$.

Proof (ii). Let X - continuous RV.

$P(X\geq 0) = 1 \Rightarrow P(X<0) = 0$ so $f_X(x) = 0$ for almost all $x<0$

Hence,

$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^{\infty} x f_X(x)\, dx \geq 0$ since the integrand ≥ 0.

Proof (iii). $X \geq Y$ a.s. $\Rightarrow X - Y \geq 0$ a.s

$\Rightarrow E(X-Y) \geq 0$, (ii)

$\Rightarrow E(X) - E(Y) \geq 0$, linearity

2.3 : Moments

Def: For each $n \in \mathbb{N}$, the n-th moment of X, or $F_X(x)$ is $\boxed{m'_n := E(X^n)}$

The n-th central moment of X is $\boxed{m_n := E(X - E(X))^n}$

$Y = X - E(X)$

$E(Y) = E(X - E(X))$

$= E(X) - E^2(X)$?

Remark: Let $n=2$. The second central moment is the variance, i.e.

$$\boxed{\begin{aligned} var(X) &:= E(X - E(X))^2 \\ SD(X) &:= \sqrt{var(X)} \end{aligned}}$$

notes: · $E(X)$ captures the location (center) of X dist.
· $var(X)$ and/or $SD(X)$ captures the spread

Consider the optimization problem $\min_{b\in\mathbb{R}} E(X-b)^2$

$E(X-b)^2 = E[X^2 - 2bX + b^2]$

$= E(X^2) - 2bE(X) + b^2$

$= b^2 - 2bE(X) + [E(X)]^2 + \underbrace{E(X^2) - (E(X))^2}_{var(X)}$

$= (b - E(X))^2 + \underbrace{var(X)}_{\geq 0}$

Proof. $var(X) = E(X - E(X))^2$

$= E(X^2 - 2XE(X) + (E(X))^2)$

$= E(X^2) - 2E^2(X) + E^2(X)$, $E(XE(X)) = E(X)E(X) = E^2(X)$

$= E(X^2) - E^2(X)$. ■

The 'location b' which solves the least squares problem is $b = E(X)$

$\Rightarrow \min_b E(X-b)^2 = var(X)$

EX (Exponential Dist.): $X \sim \text{EXP}(\lambda)$ has $E(X) = \lambda$

$var(X) = E(X^2) - E^2(X)$

$= \int_0^{\infty} x^2 \frac{1}{\lambda} e^{-x/\lambda}\, dx - \lambda^2$

$= 2\lambda^2 - \lambda^2$, IBP

$= \lambda^2.$

Thm. If X is a RV w/ finite variance, then for any $a, b \in \mathbb{R}$:

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

Recall: $E(aX+b) = aE(X)+b$

Proof. $Var(aX+b) = E[(aX+b)^2 - (aEX+b)^2]$

$= E[a^2x^2 + 2axb + b^2 - a^2(E(X))^2 - 2abE(X) - b^2]$

$= a^2 E(X^2) + 2ab E(X) - a^2 E^2(X) - 2abE(X)$, by L.O.E

$= a^2(E(X^2) - E^2(X))$

$= a^2 Var(X)$. □

Ex(Binomial dist.)

Fast way

⭑ Finish this from end-of-Tuesday class that I missed

Moment-Generating Functions

Def: Let X-RV w/ cdf F_X. The moment-generating function (mgf) is $\boxed{M_X(t) := E(e^{tX})}$

provided that the expectation exists in some neighborhood of $t=0$ (i.e. $t \in$ some $(-h,h)$ for $h>0$).

Remark: $\cdot\ M_X(t) = \begin{cases} \int e^{tx} f_X(x)\,dx & , X\text{-cont. RV} \\ \sum_x e^{tx} f_X(x) & , X\text{-discrete RV} \end{cases}$

$\cdot$ the mgf is/can be used to "characterize" the dist. (of the RV)

Thm. $\forall n \in \mathbb{N}$, $\boxed{E(X^n) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}}$

Remarks: $\cdot$ we also write RHS as $M_X^{(n)}(0)$

(n-th derivative of the mgf evaluated @ $t=0$ gives the n-th moment)

$\cdot$ gives meaning of "moment generating"

Proof (⭑). $n=0$ $\quad E(X^0) = E(1) = 1$

$M_X(0) = Ee^{0 \cdot X} = E(1) = 1.$

$n=1$ $\quad \frac{d}{dt} M_X(t) = \frac{d}{dt}\int e^{tx} f_X(x)\,dx$ or Σ

$= \int \frac{d}{dt} e^{tx} f_X(x)\,dx$

$= \int x e^{tx} f_X(x)\,dx$

$\Rightarrow \frac{d}{dt} M_X(t)\big|_{t=0} = \int x f_X(x)\,dx = E(X).$

$n=2$ $\quad \frac{d^2}{dt^2} M_X(t) = \int \frac{d}{dt} x e^{tx} f_X(x)\,dx$

$= \int x^2 e^{tx} f_X(x)\,dx$

$\Rightarrow \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = E(X^2)$

$\Downarrow$

$\vdots$

$\Downarrow$

and so on for any $n \in \mathbb{N}$. □

Ex (Gamma): For any $\alpha, \beta > 0$ define Gamma function

$$\Gamma(\alpha) = \frac{1}{\beta^\alpha}\int_0^\infty x^{\alpha-1} e^{-x/\beta}\,dx \quad (⭑)$$

Note that $\Gamma(\alpha)$ does not depend on β, b/c by change of variable,

$y = \frac{x}{\beta}$, $\frac{1}{\beta^\alpha}\int_0^\infty x^{\alpha-1} e^{-x/\beta}\,dx = \frac{1}{\beta^\alpha}\int_0^\infty (\beta y)^{\alpha-1} e^{-y} \beta\,dy$

$= \int_0^\infty y^{\alpha-1} e^{-y}\,dy$

which now doesn't depend on β, and usually is taken as the def. of $\Gamma(\alpha)$.

Define $f(x) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $x \in (0, \infty)$

f is a pdf on $(0,\infty)$ and is indeed the pdf for Gamma distribution, denoted Gamma(α, β).

⭑ alternative / equivalent expression

For $X \sim$ Gamma(α, β):

$M_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta}\,dx$

$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1} e^{-x\underbrace{(\frac{1}{\beta}-t)}_{\frac{1-\beta t}{t}}}\,dx$

⭑why? $= \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha$, if $1-\beta t > 0$

$= \left(\frac{1}{1-\beta t}\right)^\alpha$, if $t < \frac{1}{\beta}$

if $t \geq \frac{1}{\beta}$ then $\frac{1}{\beta} - t \leq 0$, then $\int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)}\,dx = M_X(t) = +\infty$

$-\beta$? via chain rule

Now, $E(X) = \frac{d}{dt} M_X(t)\Big|_{t=0} = \alpha(1-\beta t)^{-\alpha-1}\beta\Big|_{t=0}$

$= \alpha\beta$

Ex (Binomial): Let $X \sim$ Binomial(n, p)

$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$

Then,

$M_X(t) = E(e^{tX})$

$= \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x (1-p)^{n-x}$

$= \sum_{x=0}^{n} \binom{n}{x}(e^t p)^x (1-p)^{n-x}$

$= (e^t p + 1 - p)^n$, binomial formula?

Differentiate $M_X(t)$ up to the k^{th} order and evaluate it at $t=0$ to obtain $E(X^k)$

Remark: • the mgf uniquely determines all moments

Q: Does the (infinite) set of moments uniquely determine (identify) the dist?

A: not always

2.3: Moments (cont.)

RV X is defined by cdf $\boxed{F_X(x)}$ • harder to estimate

$$\begin{tikzcd} & \boxed{F_X(x)} \arrow[dl] \arrow[d] & \\ \text{mgf } M_X(t) \arrow[r, "\text{generate}"] & \text{all moments} & \end{tikzcd}$$

• easier to estimate

Q1: Does the mgf uniquely determine the cdf?

Q2: Does the set of moments $\{E(X^r)\}$ uniquely determine the cdf?

Thm. Let $F_X(x)$ and $F_Y(y)$ be two cdf's, all of whose moments exist.

i) If X and Y have bounded support, i.e. $\mathcal{X} = \{x : f_X(x) > 0\}$, then

$$\boxed{F_X(u) = F_Y(u) \ \forall u \iff E(X^r) = E(Y^r) \ \forall r = 0, 1, 2, \dots}$$

ii) If $M_X(t)$ and $M_Y(t)$ exist, and $M_X(t) = M_Y(t)$ for all t in some neighbourhood of 0, then $F_X(u) = F_Y(u) \ \forall u$

Q: What if X and Y have unbounded support? Let $X \sim f_1$ and $Y \sim f_2$

Let $Z \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$.

Suppose $X = e^Z \Rightarrow X \sim f_1(x) = \frac{1}{\sqrt{2\pi}\, x} e^{-\frac{1}{2}(\log(x))^2}$, $x \in [0, \infty)$ where X: log-normal

and $f_2(x) = f_1(x) \cdot (1 + \sin(2\pi \log(x)))$

Then, it can be checked that

$$E(X^r) = E(Y^r) = e^{r^2/2} \text{ for } r = 0, 1, 2, \dots$$

* the moments do not capture all info abt the dist. in the unbounded support scenario

Q: why do we care?

A: lies in need to approximate RV's

note: Henceforth, $M_X(t) = M_X(t)$.

Thm (convergence of MGF's leads to convergence of CDF).
Suppose $X_1, X_2, \dots$ is a sequence of random variables, each w/ mgf $M_{X_i}(t)$.

Suppose
$$\lim_{i \to \infty} M_{X_i}(t) \to M_X(t)$$
for all t in a neighbourhood of 0, and $M_X(t)$ a mgf. (*) → (*) $\exists h > 0$ s.t. $\forall t \in (-h, h)$, $M_{X_i}(t) = M_X(t)$

Then $X_i \to X$ in distribution

where X is a RV w/ mgf $M_X(t)$

i.e. $F_{X_i}(x) \to F_X(x)$ at all points x where cdf F_X is continuous

Proof (idea). Beyond scope of class

Def: $X_i \to X$ in distribution ($X_i \xrightarrow{d} X$) if $F_{X_i}(x) \to F_X(x)$ at all continuity points of F_X

• implicit in above thm.