Expectation or Mean

Defn (a) Let X be a discrete r.v with p.m.f $f(\cdot)$ and suppor S. We say that the expected value of X (or the mean of X, which we denote by $E(X)$ and defined as

$$E(X) = \sum_{x \in S} x f_X(x), \quad \text{provided}$$

$$\sum_{x \in S} |x| f_X(x)\, dx < \infty$$

(b) Let X be a continuous r.v. with pdf $f_X(x)$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx, \quad \text{provided } \int_{-\infty}^{\infty} |x| f_X(x) dx$$

is finite.

Let $h(x) : \mathbb{R} \to \mathbb{R}$. Then $E(h(X))$ is defined as

$$E(h(X)) = \sum_{x \in S} h(x) f_X(x) \quad (\text{for } X \text{ is discrete}$$

provided $\sum_{x \in S} |h(x)| f_X(x) < \infty$

For X is con continuous

$$E\,h(X) = \int_{-\infty}^{\infty} h(x)\, f_X(x)\, dx \,, \quad \text{provided } \int_{-\infty}^{\infty} |h(x)|\, f_X(x)\, dx$$

is finite.

Ex: Let X be a r.v. with p.m.f

$$f_X(x) = \begin{cases} \frac{1}{6}, & x = -2, -1, 0, 1, 2, 3 \\ 0, & o/w \end{cases}$$

Find $E(X^2)$.

Solu^n: $$E(X^2) = \sum_{x \in S} x^2 f_X(x) = 4 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \cdot \frac{1}{6} + 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6}$$

$$= \frac{19}{6}.$$

Ex: X be a r.v. with p.d.f

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & o/w. \end{cases}$$

Find $E(X^3)$.

Solu^n: $$E(X^3) = \int_{-\infty}^{\infty} x^3 f_X(x)\, dx = 2\int_0^1 x^4\, dx = 2/5$$

Some special expectation :

(i) Let $g(x) = x$, $E\, g(x) = E(X) = \mu_1' =$ mean of distribution of X. It is denoted as μ.

(ii) $g(x) = x^k$, $k = 1, 2, \dots$, $E(X^k) = \mu_k'$ - k^{th} moment of X about origin

(iii) $g(x) = (x - \mu_1')^k = (x-\mu)^k$, $E(X-\mu)^k = \mu_k = k^{th}$ moment of X about its mean or k^{th} central moment.

(iv) $\mu_2 = E(X-\mu)^2 = \sigma^2 =$ variance of X

we also denote it by $Var(X)$.

$\sqrt{\mu_2} = \sqrt{E(X-\mu)^2} = \sigma$ is called standard deviation of X.

Observations

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

$$= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu^2 + \mu^2$$

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

(ii) $(x-\mu)^2 \geq 0$ we have

$$Var(x) = E(x-\mu)^2 \geq 0$$

$$\Rightarrow E(x^2) \geq (E(x))^2$$

<u>Theorem</u>: Let X be a discrete or continuous r.v. with ~~pdf~~ pmf / pdf $f(x)$ and let $h_i : \mathbb{R} \to \mathbb{R}$

$i = 1, 2, \ldots k$ be given functions.

(a) Then for real constants $c_1, c_2, \ldots c_k$

$$E\left(\sum_{i=1}^{k} c_i h_i(x)\right) = \sum_{i=1}^{k} c_i E(h_i(x))$$

provided involved expectations are finite.

(b) Let $h_1(x) \leq h_2(x)$ $\forall x \in S$ then

$$E(h_1(x)) \leq E(h_2(x))$$

provided expectations are finite. In particulare if $E(x)$ is finite and $P(a \leq x \leq b) = 1$ for som real constant a and b ~~...~~ then $a \leq E(x) \leq b$.

(c) For any real constants a and b

$$E(ax+b) = a\,E(x) + b.$$

provided excepltation exists.

Theorem: Let X be a r.v s.t $E|X|^s < \infty$, for some $s > 0$. Then $E(|X|^r) < \infty \; \forall \; 0 < r < s$.

Moment generating Function:

Let X be a random variable with d.f. F and pdf/pmf $f(x)$.

Def^n: We say that the moment generating function (m.g.f) of X (denoted by $M_X(t)$) exists and equals

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

provided $E(e^{tX})$ is finite in $(-h, h)$ for some $h > 0$.

Remark (1) $M_X(0) = 1$. Thus $A = \{t \in \mathbb{R} : E(e^{tX}) \text{ is finite}\} \neq \phi$.

(III) Suppose that $M_X(t)$ exists is finite on $(-h, h)$ for some $h > 0$. For real constants c and d $Y = cX + d$, the m.g.f of Y also exists and is finite on $\left(-\frac{h}{|c|}, \frac{h}{|c|}\right)$ &

$$M_Y(t) = E\left(e^{t(cX+d)}\right) = e^{td} M_X(ct).$$

Theorem: Let X be a r.v. with m.g.f M_X that is finite on $(-h, h)$, $h>0$. Then

ⓐ For each $x \in \{1, 2, \dots\}$, $\mu'_k = E(X^k)$ is finite.

ⓑ For each $k \in \{1, 2, \dots\}$, $\mu'_k = E(X^k) = M_X^{(k)}(0)$,

where
$$M_X^{(k)}(0) = \left[\frac{d^k}{dt^k} M_X(t)\right]_{t=0}$$

(c)
$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu'_k, \quad t \in (-h, h), \text{ so that}$$

μ'_k is equal to coefficient of $\frac{t^k}{\underline{|k}}$ $(k=1, 2, \dots)$ around in the Maclaurin's series expansion of $M_X(t)$ around $t=0$.

Proof: Beyond the Scope.

Example: Under the notation and assumption of the above theorem define $\psi_X(t) = \ln M_X(t)$, $t \in (-h, h)$.

Then
$$\mu'_1 = \mu = E(X) = \psi_X^{(1)}(0)$$
$$\mu_2 = \sigma^2 = Var(X) = \psi_X^{(2)}(0)$$

Solu^n: $\psi_X^{(1)}(t) = \dfrac{M_X^{(1)}(t)}{M_X(t)}$, $\psi_X^{(2)}(t) = \dfrac{M_X(t)\, M_X^{(2)}(t) - \left(M_X^{(1)}(t)\right)^2}{\left(M_X(t)\right)^2}$

$$\Rightarrow \psi_X^{(1)}(0) = M_X^{(1)}(0) = E(X) \quad (\because M_X(0)=1)$$

$$\psi_X^{(2)}(0) = M_X^{(2)}(0) - \left(M_X^{(2)}(0)\right)^2$$

$$= E(X^2) - \left(E(X)\right)^2 = Var(X).$$

Example: Let X be a discrete r.v. with p.m.f.

$$f_X(x) = P_r(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0,1,2,\dots \\ 0, & o/w. \end{cases}$$

Where $\lambda > 0$. Show that the m.g.f. of X exists and is finite on whole $\mathbb{R}$. Find $M_X(t)$, mean, variance of X and $E(X^3)$.

solu^n: we have X is a discrete r.v. with p.m.f.

$$f_X(x) = P_r(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0,1,2,\dots \\ 0 & o/w \end{cases}$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}, \quad \forall\, t \in \mathbb{R}$$

Thus m.g.f of X exists and finite on whole of $\mathbb{R}$.

$$M_X(t) = e^{\lambda(e^t-1)}, \quad t \in \mathbb{R}$$

Now $\psi_X(t) = \ln(M_X(t)) = \lambda(e^t-1)$

$$\psi_X^{(1)}(t) = \lambda e^t, \quad \psi_X^{(2)}(t) = \lambda e^t \quad \forall t \in \mathbb{R}$$

$$E(X) = \psi_X^{(1)}(0) = \lambda \quad \& \quad Var(X) = \psi_X^{(2)} = \lambda.$$

Again

$$M_X^{(1)}(t) = \lambda e^t e^{\lambda(e^t-1)} = \lambda e^t M_X(t)$$

$$M_X^{(2)}(t) = \lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t)$$

$$M_X^{(3)}(t) = \lambda e^t M_X^{(2)}(t) + 2\lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t)$$

$$\Rightarrow M_X^{(1)}(0) = E(X) = \lambda$$

$$E(X^2) = M_X^{(2)}(0) = \lambda^2 + \lambda.$$

$$E(X^3) = M_X^{(3)}(0) = \lambda^3 + 3\lambda^2 + \lambda.$$

Ex: Let X be a continuous r.v. with p.d.f

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & o/w \end{cases}, \quad \text{where } \lambda > 0.$$

Find m.g.f. of X, mean, variance of X. (H.W.)

Defn: (Equality in Distribution): Let X and Y be two r.v's with d.f's F_X and F_Y respectively. We say that X and Y have the same distn written as $X \overset{d}{=} Y$ if

$$F_X(x) = F_Y(x) \quad \forall\, x \in \mathbb{R}.$$

Remark (i) Let X and Y be discrete r.v's with p.m.f's $f_X(x)$ and $f_Y(y)$ respectively. ~~We say that~~ Then

$$X \overset{d}{=} Y \iff f_X(x) = f_Y(x) \quad \forall\, x \in \mathbb{R}.$$

(ii) Let X and Y be continuous r.v's. Then $X \overset{d}{=} Y$ iff there exists versions of p.d.f's f_X and f_Y respectively such that $f_X(x) = f_Y(x) \;\forall\, x \in \mathbb{R}$.

(iii) Suppose $X \overset{d}{=} Y$. Then for any $h: \mathbb{R} \to \mathbb{R}$ $h(X) \overset{d}{=} h(Y)$ and hence $E(h(X)) = E(h(Y))$.

Theorem: Let X and Y be r.v's such that for some $h>0$ $M_X(t) = M_Y(t) \;\forall\; t \in (-h, h)$. Then $X \overset{d}{=} Y$.

Example: For any $p \in (0,1)$ and positive integer n let $X_{p,n}$ be a discrete r.v. with p.m.f.

$$f_{p,n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x \in \{0,1,2,\dots,n\},\ n \in \mathbb{N} \\ 0 & \text{o/w} \end{cases}$$

Define $Y_{p,n} = n - X_{p,n}$, $p \in (0,1)$, $n \in \mathbb{N}$.

Using the m.g.f. of $X_{p,n}$ show that $Y_{p,n} \overset{d}{=} X_{1-p,n}$

Find $E(X_{\frac{1}{2},n})$.

Soln: We have

$$M_{X_{p,n}}(t) = E\left(e^{tX_{p,n}}\right) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

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

$t \in \mathbb{R}$.

$$M_{Y_{p,n}}(t) = E\left(e^{tY_{p,n}}\right) = E\left(e^{t(n-X_{p,n})}\right)$$

$$= e^{nt} M_{X_{p,n}}(-t) = e^{nt}\left(1-p+pe^{-t}\right)^n$$

$$= \left(p + (1-p)e^t\right)^n = \left(1 - (1-p) + (1-p)e^t\right)^n$$

$$= M_{X_{1-p,n}}(t) \quad \forall\ t.$$

$$\Rightarrow Y_{n,p} \overset{d}{=} X_{1-p,n}.$$

Now for $p = \frac{1}{2}$, so $X_{\frac{1}{2},n} \stackrel{d}{=} n - X_{\frac{1}{2},n}$

$$E\left(X_{\frac{1}{2},n}\right) = E\left(n - X_{\frac{1}{2},n}\right) \Rightarrow E\left(X_{\frac{1}{2},n}\right) = \frac{n}{2}$$

Some Inequalities :

Theorem : Let X be a r.v and let $g: \mathbb{R} \to \mathbb{R}$ be a non-negative function such that $E(g(x))$ is finite. Then for any $c > 0$

$$P_r\left(g(x) \geq c\right) \leq \frac{E(g(x))}{c}$$

Proof: We will prove it for the case of continuous r.v.

Let $A = \{x \in \mathbb{R} : g(x) \geq c\}$. Let $f_X(x)$ denote the p.d.f of X. Then

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx = \int_A g(x) f_X(x)\, dx + \int_{A^c} g(x) f_X(x)\, dx$$

$$\geq \int_A g(x) f_X(x)\, dx \geq c \int_A f_X(x)\, dx \quad \left[\because g(x) \geq c \text{ in } A\right]$$

$$= c\, P(g(x) \geq c)$$

$$\Rightarrow P(g(x) \geq c) \leq \frac{E(g(x))}{c}.$$

Corollary: (a) Let $g : [0,\infty) \to \mathbb{R}$ be a nonnegative and strictly increasing fun^n such that $E(g(x))$ is finite.
Let any $\epsilon > 0$ s.t. $g(c) > 0$,

$$P_r(|x| \geq c) \leq \frac{E(g(|x|))}{g(c)}$$

(b) Let $r > 0$ and $t > 0$. Then

$$P_r(|x| > t) \leq \frac{E(|x|^r)}{t^r} \quad \text{(Markov's Inequality)}$$

Proof: (a)

$$P_r(|X| \geq c) = P_r(g(|X|) \geq g(c)) \quad \left(\text{Since } g \text{ is strictly } \uparrow\right)$$

$$\leq \frac{E(g(|X|))}{g(c)} \quad \left(\text{by previous theorem}\right)$$

(b) We take $g(x) = x^r$, $x \geq 0$, $r > 0$

Then g is strictly increasing on $[0,\infty)$ and is non-negative

$$P(|X| \geq t) \leq \frac{E(g(|X|))}{g(t)} = \frac{E(|X|^r)}{t^r}$$

This proves the result.

Theorem: (Chebyshev Inequality): Let X be a r.v. with finite ~~[illegible]~~ variance σ^2 and $E(X)=\mu$. Then for any $\epsilon > 0$

$$\Pr\left(|X-\mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}$$

Using corollary take $Y = |X-\mu|$, $r=2$

Then $$\Pr\left(|Y| \geq \epsilon\right) \leq \frac{E(|Y|^2)}{\epsilon^2} = \frac{E(|X-\mu|^2)}{\epsilon^2}$$

$$\Rightarrow \Pr\left(|X-\mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}$$

Def^n: Let $-\infty \leq a < b \leq \infty$. A function $\psi : (a,b) \to \mathbb{R}$ is said to be a convex function if

$$\psi(\alpha x + (1-\alpha) y) \leq \alpha \psi(x) + (1-\alpha)\psi(y) \quad \forall$$
$$x, y \in (a,b) \text{ and } \alpha \in (0,1).$$

The function $\psi(\cdot)$ is said to be strictly convex if the above inequality is strict.

We state the following theorem without proof.

Theorem: (i) Let $\psi : (a,b) \to \mathbb{R}$ be a convex fu^n. Then ψ is continuous on (a,b), and is almost everywhere diff^n (i.e. if D is the set of points where ψ is not differentiable then D does not contain interval,

(ii) Let $\psi : (a,b) \to \mathbb{R}$ be a differentiable fun. Then ψ is convex (strictly convex) on (a,b) iff ψ' is non-decreasing (strictly increasing) on (a,b).

(iii) Let $\psi : (a,b) \to \mathbb{R}$ be a twice differentiable fun. Then ψ is convex (strictly convex) on (a,b) iff $\psi''(x) \geq (>0) = 0 \quad \forall\, x \in (a,b)$.

Theorem: (Jensen's Inequality): Let $\psi : (a,b) \to \mathbb{R}$ be a convex fun and let X be a r.v. with support $S \subseteq (a,b)$. Then

$$E(\psi(X)) \geq \psi(E(X))$$

provided the expectations exists.

Proof: We give the proof for the special case where ψ is twice differentiable on (a,b) so that $\psi''(x) \geq 0$ $\forall\, x \in (a,b)$. Let $\mu = E(X)$. Expand $\psi(x)$ into a Taylor series about μ we get

$$\psi(x) = \psi(\mu) + (x-\mu)\,\psi'(\mu) + \frac{(x-\mu)^2}{2!}\,\psi''(\xi) \quad \forall\, x \in (a,b)$$

for some ξ between μ and x.

$$\Rightarrow \psi(x) \geq \psi(\mu) + (x-\mu)\,\psi'(\mu)$$

$$\Rightarrow E(\psi(x)) \geq E(\psi(\mu) + (x-\mu)\,\psi'(\mu)) = \psi(\mu) = \psi(E(X)).$$

Example: For any r.v. X, (i) $E(X^2) \geq (E(X))^2$

(ii) $E(|X|) \geq |E(X)|$ (iii) For a random variable $X > 0$ $E(\ln X) \leq \ln E(X)$

Solun: (i) Take $\psi(x) = x^2 \to$ convex

(ii) Take $\psi(x) = |x| \to$ convex.

(iii) Take $\psi(x) = -\ln x$ – convex on $(0, \infty)$.

Apply Jensen's Inequality:

Symmetric Distribution: A r.v. X is symmetric about a point α if $P(X \geq \alpha + x) = P(X \leq \alpha - x)$ $\forall x$.

or

$$F(\alpha - x) = 1 - F(\alpha + x) + P(X = \alpha + x).$$

If $\alpha = 0$, i.e. symmetric about point 0.

$$F(-x) = 1 - F(x) + P(X = x)$$

If X is continuous then

$$F(-x) = 1 - F(x) \quad \& \quad \text{for density}$$

$$f_X(-x) = f_X(x) \ \forall x \in \mathbb{R}$$

EX: $P(X=-1) = \frac{1}{4}$, $P(X=0) = \frac{1}{2}$, $P(X=1) = 1/4$

Symmetric about 0.

Some examples

Ex (1) $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty$

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1+x^2}\, dx = \frac{1}{\pi} \tan^{-1} x \Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2}+\frac{\pi}{2}\right) = 1.$$

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x}{1+x^2}\, dx$$

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

$= \lim_{a\to\infty} \frac{1}{\pi} \log(1+x^2)\Big|_0^a$ which divergent.

So the expectation does not exists.

Ex (2) consider a prob. dist^n as

$$P\left(X = (-1)^{j+1} \frac{3^j}{j!}\right) = \frac{2}{3^j}, \quad j = 1, 2, \ldots$$

$\sum_{j=1}^{\infty} |x_j|\, P(X = x_j) = \sum_{j=1}^{\infty} \frac{3^j}{j} \cdot \frac{2}{3^j}$ is divergent

so $E(X)$ does not exists

Example 3 A package of 4 bulbs contain one defective The bulbs are tested one by one with out replacement until the defective is detected Find the prob dist^n of number of testing required.

Solu^n.

$X \rightarrow$ no. of testing required.

$\rightarrow 1, 2, 3$

$P(X=1) = \frac{1}{4}$, $P(X=2) = \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4}$

$P(X=3) = \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}$

$\parallel$

P(Third bulb is defective) + P(Third bulb is not defective).

Ex. 4 $f_X(x) = \begin{cases} \frac{2}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$

$E(X) = \int_1^\infty \frac{2x}{x^3} dx = 2$, $E(X^2) = \int_1^\infty \frac{2x^2}{x^3}$ not exists

So Here $E(X)$ is exists but $E(X^2)$ does not exists.

Theorem: Let X be a r.v s.t. $E|X|^s < \infty$ for some $s > 0$. Then $E(|X|^r) < \infty$ $\forall\ 0 < r < s$. i.e. if the moment of order s exists then the moment of order r $(0 < r < s)$ exists. for a given r.v. X.

Proof: Let X be a continuous r.v. with pdf $f_X(x)$

$$E|X|^r = \int_{-\infty}^{\infty} |x|^r f_X(x)\, dx$$

$$= \int_{|x|\le 1} |x|^r f_X(x)\, dx + \int_{|x|>1} |x|^r f_X(x)$$

$$\le \int_{|x|\le 1} f_X(x)\, dx + \int_{|x|>1} |x|^s f_X(x)\, dx$$

$$\le P(|X| \le 1) + \int_{-\infty}^{\infty} |x|^s f_X(x)\, dx < \infty.$$

$$\because P(|X| \le 1) \le 1 \ \&\ E|X|^s < \infty.$$

This proves the result.

Def^n: (Degenerate Disf^n): A random variable X is said to be degenerate at a point $c \in \mathbb{R}$ if $P(X=c) = 1$.

Suppose that a r.v. X is degenerate at $c \in \mathbb{R}$. Then clearly X is of discrete type with support $S=\{c\}$ and disf^n fu^n is

$$F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$

and p.m.f

$$f_X(x) = \begin{cases} 1, & x = c \\ 0, & o/w \end{cases}$$

$$E(X) = c, \qquad Var(X) = 0.$$

Ex: The number of customers who visit a store every day is a r.v. X with $\mu = 18$ and $\sigma = 2.5$. With what prob can be assert that there will be between 8 to 28 customes.?

$$P(8 \leq X \leq 28) = P(-10 < X - 18 < 10)$$

$$= P(|X-18| \leq 0) \geq 1 - \frac{\sigma^2}{100} = 1 - \frac{6.25}{100}$$

$$= \frac{15}{16}.$$