IC105: Probability and Statistics **2021-22-M**

Lecture 10: Some Inequalities

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

Example 10.1. *(a) Let X be a discrete r.v. with p.m.f.*

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots, \\ 0, & \textit{otherwise}, \end{cases}$$

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 $E(X^3)$.

(b) Let X be a continuous r.v. with p.d.f.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \textit{otherwise}, \end{cases}$$

where $\lambda > 0$. Find m.g.f., mean, variance of X and $E(X^r)$, $r = 1, 2, \ldots$ (provided they exist).

(c) Let X be a continuous r.v. having the p.d.f. $f(x) = \dfrac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$ (called Cauchy p.d.f. and corresponding probability distribution is called Cauchy distribution). Show that the m.g.f. of X does not exist.

Solution: (a) We have

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}, \ \forall\, t \in \mathbb{R}.$$

Thus, m.g.f. of X exists and finite on whole of $\mathbb{R}$ and $M_X(t) = e^{\lambda(e^t-1)}, \ t \in \mathbb{R}$.

Now $\psi_X(t) = \ln(M_X(t)) = \lambda(e^t - 1) \implies \psi_X^{(1)}(t) = \lambda e^t = \psi_X^{(2)}(t), \ \forall\, t \in \mathbb{R}$.

Thus, $E(X) = \psi_X^{(1)}(0) = \lambda$ and $\mathrm{Var}(X) = \psi_X^{(2)}(0) = \lambda$. Again,

$$\begin{aligned}
&M_X^{(1)}(t) = \lambda e^t e^{\lambda(e^t-1)} = \lambda e^t M_X(t) \implies M_X^{(1)}(0) = E(X) = \lambda,\\
&M_X^{(2)}(t) = \lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t) \implies M_X^{(2)}(0) = E(X^2) = \lambda^2 + \lambda,\\
&M_X^{(3)}(t) = \lambda e^t M_X^{(2)}(t) + 2\lambda e^t M_X^{(1)}(t) + \lambda e^t M_X(t) \implies M_X^{(3)}(0) = E(X^3) = \lambda^3 + 3\lambda^2 + \lambda.
\end{aligned}$$

Alternatively, for $t \in \mathbb{R}$,

$$\begin{aligned}
M_X(t) &= e^{\lambda(e^t-1)}\\
&= 1 + \lambda(e^t - 1) + \frac{\lambda^2(e^t-1)^2}{2!} + \frac{\lambda^3(e^t-1)^3}{3!} + \cdots\\
&= 1 + \lambda\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right) + \frac{\lambda^2}{2!}\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right)^2 + \frac{\lambda^3}{3!}\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right)^3 + \cdots\\
&= 1 + \lambda t + t^2\left(\frac{\lambda}{2!} + \frac{\lambda^2}{2!}\right) + t^3\left(\frac{\lambda}{3!} + \frac{2\lambda^2}{(2!)^2} + \frac{\lambda^3}{3!}\right) + \cdots
\end{aligned}$$

Thus,

$$E(X) = \text{coefficient of } t \text{ in the expansion of } M_X(t) = \lambda,$$
$$E(X^2) = \text{coefficient of } \frac{t^2}{2!} \text{ in the expansion of } M_X(t) = \lambda^2 + \lambda,$$
$$E(X^3) = \text{coefficient of } \frac{t^3}{3!} \text{ in the expansion of } M_X(t) = \lambda^3 + 3\lambda^2 + \lambda.$$

(b) $\int_{-\infty}^{\infty} e^{tx} f_X(x)\mathrm{d}x = \lambda \int_{-\infty}^{\infty} e^{-\lambda(1-t/\lambda)x}\mathrm{d}x < \infty$, if $t < \lambda$. Thus the m.g.f. of X exists and, for $t < \lambda$,

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} = 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \cdots + \frac{t^r}{\lambda^r} + \cdots.$$

For $r = 1, 2, \ldots$

$$\mu_r' = E(X^r) = \text{coefficient of } \frac{t^r}{r!} \text{ in the expansion of } M_X(t) = \frac{r!}{\lambda^r}, \;\; r \in \{1, 2, \ldots\}.$$

Alternatively,

$$M_X^{(1)}(t) = \frac{1}{\lambda}\left(1 - \frac{t}{\lambda}\right)^{-2}, \; M_X^{(2)}(t) = \frac{2}{\lambda^2}\left(1 - \frac{t}{\lambda}\right)^{-3} \text{ and } M_X^{(r)}(t) = \frac{r!}{\lambda^r}\left(1 - \frac{t}{\lambda}\right)^{-(r+1)}, \;\; t < \lambda.$$

This implies

$$E(X^r) = M_X^{(r)}(0) = \frac{r!}{\lambda^r}, \;\; r = 1, 2, \ldots \text{ and } \; \mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(c) Since $E(X)$ is not finite, the m.g.f. of X does not exist.

Definition 10.2 (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 distribution (written as $X \overset{d}{=} Y$) if $F_X(x) = F_Y(x), \; \forall\, x \in \mathbb{R}$.*

Remark 10.3. *(i) Let X and Y be two discrete r.v.'s with p.m.f.'s f_X and f_Y, respectively. Then,*

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

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

(iii) Suppose $X \overset{d}{=} Y$, then for any Borel measurable function $h : \mathbb{R} \to \mathbb{R}, \;\; h(X) \overset{d}{=} h(Y)$ and hence $E(h(X)) = E(h(Y))$.

Theorem 10.4. *Let X and Y be r.v.'s such that for some $c > 0, \;\; M_X(t) = M_Y(t), \; \forall\, t \in (-c, c)$. Then, $X \overset{d}{=} Y$.*

Proof. Special Case: Suppose that X and Y are discrete r.v.'s with support $S_X = S_Y = \{1, 2, \ldots\}$, $p_k = P(X = k)$ and $q_k = P(Y = k)$, $k = 1, 2, \ldots$. Then

$$\begin{aligned}
& M_X(t) = M_Y(t), \; \forall\, t \in (-c, c), \text{for some } c > 0 \\
\implies & \sum_{k=1}^{\infty} e^{kt} p_k = \sum_{k=1}^{\infty} e^{kt} q_k, \;\; \forall\, t \in (-c, c) \\
\implies & \sum_{k=1}^{\infty} \Lambda^k p_k = \sum_{k=1}^{\infty} \Lambda^k q_k, \;\; \forall\, \Lambda \in (e^{-c}, e^c) \\
\implies & p_k = q_k, \;\; \forall\, k = 1, 2, \ldots,
\end{aligned}$$

since if two power series are equal over an interval then their coefficients are the same. Thus, $X \overset{d}{=} Y$. □

Example 10.5. *For any $p \in (0,1)$ and positive integer n, let $X_{p,n}$ be a discrete r.v. with p.m.f.*

$$f_{p,n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \textit{if } x = \{0, 1, \ldots, n\}, \\ 0, & \textit{otherwise}. \end{cases}$$

Here, $p \in (0,1)$ and $n \in \mathbb{N}$. (Such a r.v. or probability distribution is called binomial r.v. or distribution with n trials and probability of success p). Define $Y_{p,n} = n - X_{p,n}$. Using the m.g.f. of $X_{p,n}$, show that $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$. Find $E(X_{1/2,n})$.

Solution: We have

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

Now

$$\begin{aligned} M_{Y_{p,n}}(t) = E\left(e^{tY_{p,n}}\right) &= E\left(e^{t(n - X_{p,n})}\right) \\ &= e^{nt} M_{X_{p,n}}(-t) = e^{nt}(1 - p + pe^{-t})^n \\ &= (p + (1-p)e^t)^n = (1 - (1-p) + (1-p)e^t)^n = M_{X_{1-p,n}}(t), \ \ \forall\, t \in \mathbb{R}. \end{aligned}$$

Thus, $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$.

Alternatively,

$$\begin{aligned} f_{Y_{p,n}}(y) &= P(Y_{p,n} = y) \\ &= P(X_{p,n} = n - y) \\ &= \begin{cases} \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)}, & \text{if } n - y = \{0, 1, \ldots, n\}, \\ 0, & \text{otherwise}. \end{cases} \\ &= \begin{cases} \binom{n}{y} (1-p)^y (1 - (1-p))^{n-y}, & \text{if } y = \{0, 1, \ldots, n\}, \\ 0, & \text{otherwise}. \end{cases} \\ &= f_{X_{1-p,n}}(y), \ \ \forall\, y \in \mathbb{R}. \end{aligned}$$

Thus, $Y_{p,n} \stackrel{d}{=} X_{1-p,n}$.

Now for $p = 1/2$, $X_{1/2,n} \stackrel{d}{=} n - X_{1/2,n}$. Thus, $E(X_{1/2,n}) = E(n - X_{1/2,n}) \implies E(X_{1/2,n}) = n/2$.

Example 10.6. *Let X be a r.v. with p.d.f. $f_X(x) = \frac{e^{-|x|}}{2}$, $-\infty < x < \infty$ and let $Y = -X$. Show that $Y \stackrel{d}{=} X$ and hence show that $E(X) = 0$.*

Solution: We have

$$M_Y(t) = E(e^{tY}) = E(e^{-tX}) = \int_{-\infty}^{\infty} e^{-tx} \frac{e^{-|x|}}{2} \mathrm{d}x = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-|x|}}{2} \mathrm{d}x = M_X(t), \ \ \forall\, t \in (-1, 1).$$

$$\left[M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-|x|}}{2} \mathrm{d}x = \int_{-\infty}^{0} e^{tx} \frac{e^{x}}{2} \mathrm{d}x + \int_{0}^{\infty} e^{tx} \frac{e^{-x}}{2} \mathrm{d}x \right.$$
$$= \frac{1}{2} \left(\int_0^{\infty} e^{-(1+t)x} \mathrm{d}x + \int_0^{\infty} e^{-(1-t)x} \mathrm{d}x \right)$$
$$\left. = \frac{1}{2} \left(\frac{1}{1+t} + \frac{1}{1-t} \right) = \frac{1}{1-t^2}, \;\; \forall\, t \in (-1,1) \implies X \stackrel{d}{=} Y. \right]$$

Alternatively, the p.d.f. of Y is

$$f_Y(y) = \frac{e^{-|y|/2}}{2} = f_X(y), \;\; \forall - \infty < y < \infty \implies X \stackrel{d}{=} Y.$$

Thus, $E(Y) = E(X) \implies E(-X) = E(X) \implies E(X) = 0$ (since $\int_{-\infty}^{\infty} |x| f_X(x) \mathrm{d}x < \infty$).

10.1. Inequalities

Inequalities provide estimates of probabilities when they can not be evaluated precisely.

Theorem 10.7. *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 $c > 0$,*

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

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,

$$\begin{aligned}
E(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(x) [I_A(x) + I_{A^c}(x)] f_X(x) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(x) I_A(x) f_X(x) \mathrm{d}x + \int_{-\infty}^{\infty} g(x) I_{A^c(x)} f_X(x) \mathrm{d}x \\
&\geq \int_{-\infty}^{\infty} g(x) I_A(x) f_X(x) \mathrm{d}x \\
&\geq c \int_{-\infty}^{\infty} I_A(x) f_X(x) \mathrm{d}x \\
&= c \int_A f_X(x) \mathrm{d}x = cP(g(X) \geq c) \implies P(g(X) \geq c) \leq \frac{E(g(X))}{c}.
\end{aligned}$$

This completes the proof. $\square$

Corollary 10.8. *(a) Let $g : [0, \infty) \to \mathbb{R}$ be a non-negative and strictly increasing function such that $E(g(X))$ is finite. Then, for any $c > 0$ such that $g(c) > 0$,*

$$P(|X| \geq c) \leq \frac{E(g(|X|))}{g(c)}.$$

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

$$P(|X| \geq t) \leq \frac{E(|X|^r)}{t^r}, \quad \textit{(Markov's inequality)}$$

provided $E(|X^r|) < \infty$. In particular, $P(|X| \geq t) \leq \frac{E(|X|)}{t}$, provided $E(|X|) < \infty$.

Proof. (a) Note that

$$\begin{aligned} P(|X| \geq c) &= P(g(|X|) \geq g(c)) \text{ (since } g \text{ is strictly increasing)} \\ &\leq \frac{E(g(|X|))}{g(c)} \text{ (by Theorem 10.7).} \end{aligned}$$

(b) We take $g(x) = x^r,\ \ x \geq 0,\ r > 0$. Then, g is strictly increasing on $[0, \infty)$ and is non-negative. Using (a) we get

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

This proves the result. □

Theorem 10.9 (Chebyshev Inequality)**.** *Let X be a r.v. with finite variance σ^2 and $E(X) = \mu$. Then, for any $\epsilon > 0$,*

$$P(|X - \mu| \geq \epsilon\sigma) \leq \frac{1}{\epsilon^2}.$$

Proof. Using the above Corollary

$$P(|X - \mu| \geq \epsilon\sigma) \leq \frac{E(|X - \mu|^2)}{\epsilon^2\sigma^2} = \frac{E((X - \mu)^2)}{\epsilon^2\sigma^2} = \frac{1}{\epsilon^2}.$$

This completes the proof. □

Example 10.10 (The above bounds are sharp)**.** *Let X be a r.v. with p.m.f.*

$$f(x) = \begin{cases} \frac{1}{8}, & \text{if } x = -1, 1, \\ \frac{3}{4}, & \text{if } x = 0, \\ 0, & \text{otherwise}. \end{cases}.$$

Then $E(X^2) = \frac{1}{4}$ and $P(|X| \geq 1) = \frac{1}{4}$.

Using the Markov inequality, $P(|X| \geq 1) \leq E(X^2) = \frac{1}{4}$.

Example 10.11. *Let X be a r.v. with p.d.f.*

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{if } -\sqrt{3} < x < \sqrt{3}, \\ 0, & \text{otherwise}. \end{cases}.$$

Then $\mu = E(X) = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x}{2\sqrt{3}} \mathrm{d}x = 0$, $\sigma^2 = E(X^2) = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x^2}{2\sqrt{3}} \mathrm{d}x = 1$ and

$$P(|X| \geq \frac{3}{2}) = 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}} \mathrm{d}x = 1 - \frac{\sqrt{3}}{2} = 0.134.$$

Using the Markov inequality $P(|X| \geq \frac{3}{2}) \leq \frac{4}{9} E(X^2) = \frac{4}{9} = 0.444\ldots$ (considerably conservative).

Definition 10.12. *Let $-\infty \leq a < b \leq \infty$. A function $\psi : (a, b) \to \mathbb{R}$ is said to be a convex function if*

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

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

We state the following theorem without proof.

Theorem 10.13. *(i) Let $\psi : (a, b) \to \mathbb{R}$ be a convex function. Then, ψ is continuous on (a, b) and is almost everywhere differentiable (i.e. if D is the set of points where ψ is not differentiable then D does not contain any interval).*

(ii) Let $\psi : (a, b) \to \mathbb{R}$ be a differentiable function. 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 function. Then, ψ is convex (strictly convex) on (a, b) iff

$$\psi''(x) \geq (>) 0, \ \forall\, x \in (a, b).$$

Theorem 10.14 (Jensen's Inequality)**.** *Let $\psi : (a, b) \to \mathbb{R}$ be a convex function and let X be a r.v. with d.f. F having support $S \subseteq (a, b)$. Then,*

$$E(\psi(X)) \geq \psi(E(X)), \quad \textit{provided the expectations exist.}$$

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

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

for some ξ between μ and x. Thus,

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

This completes the proof. $\square$

Example 10.15. *(a) For any r.v. X, $E(X^2) \geq (E(X))^2$ [take $\psi(x) = x^2$, $x \in \mathbb{R}$ is convex, apply Jensen's Inequality] and $E(|X|) \geq |E(X)|$ [Take $\psi(x) = |x|$, $x \in \mathbb{R}$ is convex and apply Jensen's Inequality].*

(b) For any r.v. X with $P(X > 0) = 1$, $E(\ln X) \leq \ln E(X)$ [Take $\psi(x) = -\ln x$ is convex on $(0, \infty)$ and apply Jensen's Inequality].

(c) For any r.v. X, $E(e^X) \geq e^{E(X)}$ [Take $\psi(x) = e^x$, $x \in \mathbb{R}$ is convex and apply Jensen's Inequality].

(d) For any r.v. X with $P(X > 0) = 1$, $E(X)E(1/X) \geq 1$ [Take $\psi(x) = 1/x$, $x > 0$ is convex and apply Jensen's Inequality].

Example 10.16. *Let $a_1, a_2, \ldots, a_n$, $w_1, w_2, \ldots, w_n$ be positive constants such that $\sum_{i=1}^n w_i = 1$. Prove the AM-GM-HM inequality*

$$\sum_{i=1}^n a_i w_i \geq \prod_{i=1}^n a_i^{w_i} \geq \frac{1}{\sum_{i=1}^n \frac{w_i}{a_i}}, \quad (AM \geq GM \geq HM).$$

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

$$f(x) = \begin{cases} w_i, & \text{if } x = a_i, i = 1, 2, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}.$$

Then $\psi(x) = -\ln x$, $x > 0$ is a convex function. Therefore

$$\begin{aligned}
&E(\psi(X)) \geq \psi(E(X)) \\
&\implies E(-\ln X) \geq -\ln E(X) \\
&\implies -\sum_{i=1}^{n} (\ln a_i) w_i \geq -\ln \left(\sum_{i=1}^{n} a_i w_i \right) \\
&\implies \ln \left(\sum_{i=1}^{n} a_i w_i \right) \geq \ln \left(\prod_{i=1}^{n} a_i^{w_i} \right) \implies \sum_{i=1}^{n} a_i w_i \geq \prod_{i=1}^{n} a_i^{w_i}.
\end{aligned}$$

Replacing a_i's by $\frac{1}{a_i}$'s, we get $\displaystyle\sum_{i=1}^{n} \frac{w_i}{a_i} \leq 1 / \prod_{i=1}^{n} a_i^{w_i}$. Therefore,

$$\sum_{i=1}^{n} a_i w_i \geq \prod_{i=1}^{n} a_i^{w_i} \geq \frac{1}{\sum_{i=1}^{n} \frac{w_i}{a_i}}.$$