Indian Institute of Technology, Kanpur
Probability Theory, 18–19

Instructors: Purushottam Kar, Neeraj Misra
Authors: Gurpreet Singh
Email: guggu@iitk.ac.in

SCRIBE

6

Equality and Inequalities in Distributions

1. Equality in Distribution

Definition 6.1. Two random variables X and Y are said to have the same distribution (XY) if they have the same C.D.F. *i.e.* $F_X(x) = F_Y(x) \, \forall \, x \in \mathbb{R}$

Result 6.1.1. Let X and Y be two random variables with p.m.f / p.d.f f_X and f_Y respectively. Then

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

(ii) for some $h > 0$, $M_X(t) = M_y(t), \; \forall \, t \in (-h, h) \implies X \overset{\Delta}{=} Y$

(iii) $X \overset{\Delta}{=} Y \implies h(X) \overset{\Delta}{=} h(Y)$ for any function $h : \mathbb{R} \to \mathbb{R}$

Definition 6.2 (Symmetric Distribution)**.** A random variable X is said to have a symmetric distribution about a point $\mu \in \mathbb{R}$ if $X - \mu \overset{\Delta}{=} \mu - X$

Result 6.2.1. Let X be a random variable with p.m.f / p.d.f f_X and CDF F_X. Then for some $\mu \in \mathbb{R}$

(i) If $f_X(x - \mu) = f_X(\mu - x), \; \forall \, x \in \mathbb{R}$, then the distribution of X is symmetric about μ

(ii) Distribution of X is symmetric about μ iff $\forall \, x \in \mathbb{R}$, $F_X(\mu + x) + F_X\left((\mu - x)^-\right) = 1$

(iii) If distribution of X is symmetric about μ and $\mathbb{E}\left[\, X \,\right]$ exists, then $\mathbb{E}\left[\, X \,\right] = \mu$

(iv) If distribution of X is symmetric about μ, then $F_X(\mu^-) \leq \frac{1}{2} \leq F_X(\mu)$ (Equality holds if F_X is continuous)

(v) If distribution of X is symmetric about μ, then $\mathbb{E}\left[(X - \mu)^{2m-1}\right] = 0$ where $m \in \{1, 2 \ldots\}$ provided the moments exist

2. Inequalities

Definition 6.3. A function $\phi : (a, b) \to \mathbb{R}$, where $a, b \in \mathbb{R}$ is said to be

(i) **convex** if $\phi(\alpha x + (1 - \alpha)y) \leq \alpha\phi(x) + (1 - \alpha)\phi(y)$

(ii) **concave** if $\phi(\alpha x + (1 - \alpha)y) \geq \alpha\phi(x) + (1 - \alpha)\phi(y)$

We use the concept of convexity (concavity) to derive some inqualities for prabability distributions.

Result 6.3.1. If a function $\phi : (a, b) \to \mathbb{R}$ is a convex (concave) function and is differentiable on (a, b), then ϕ' is increasing (decreasing) on (a, b).

A popular inequality derived using the nature of functions is Jensen Inequality

Result 6.3.2 (Jensen Inequality)**.** Let X be a random variable with support $S_X \subseteq (a,b)$ and let $\phi : (a,b) \to \infty$ be a function, where $a, b \in \mathbb{R}$. Then

1. if ϕ is a *convex* function,

$$\mathbb{E}[\phi(X)] \quad \geq \quad \phi(\mathbb{E}[X])$$

2. if ϕ is a *concave* function,

$$\mathbb{E}[\phi(X)] \quad \leq \quad \phi(\mathbb{E}[X])$$

2.1 Concentration (Tail) Bounds

Concentration inequalities provide bounds on how a random variable deviates from some value.

The base of many of the inequalities or bounds derived in probability theory is based on the following inequality.

Theorem 6.1. Let $g : \mathbb{R} \to [0, \infty)$ be a non-negative function such that $\mathbb{E}[g(X)] < \infty$. Then, for any $c > 0$,

$$\mathbb{P}[g(X) > c] \leq \frac{\mathbb{E}[g(X)]}{c}$$

Proof. (For absolutely continuous case) Let $A = \{x \in \mathbb{R} \mid g(x) > c\}$. Then

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \\
&= \int_A g(x) f_X(x) dx + \int_{A^c} g(x) f_X(x) dx \\
&\geq \int_A g(x) f_X(x) dx \\
&\geq c \int_A f_X(x) dx \\
&= c \cdot \mathbb{P}[g(X) > c]
\end{aligned}$$

Hence, we can say

$$\mathbb{P}[g(X) > c] \quad \leq \quad \frac{\mathbb{E}[g(X)]}{c}$$

□

Exercise 6.1. Prove the above theorem for the discrete case

Corollary 6.1.1. Let $r, c > 0$, then if $\mathbb{E}[|X|^r]$ exists

$$\mathbb{P}[|X| > c] \quad \leq \quad \frac{\mathbb{E}[|X|^r]}{c^r}$$

We can now discuss various inequalities and bounds for random variables

1. **Markov Inequality**

   Suppose that $\mathbb{E}[|X|]$ exists, then

$$\mathbb{P}[|X| > c] \quad \leq \quad \frac{\mathbb{E}[|X|]}{c}$$

2. **Chebychev Inequality**

Let X be a random variable with finite mean $\mu = \mathbb{E}[X]$ and finite variance $\sigma^2 = \mathrm{Var}(X)$. Then for any $\epsilon > 0$

$$\mathbb{P}[\,|X - \mu| > \epsilon\sigma\,] \quad \leq \quad \frac{1}{\epsilon^2}$$

3. **Chernoff's Bound**

Suppose we have N independent and identical random variables $\{X_n\}_{n\in[N]}$ such that $\forall\, n \in [N]$, $X_n \in \{0,1\}$ and $\mathbb{E}[X_n] = p$. Then for every $\epsilon \in (0,1)$

$$\begin{aligned}
\mathbb{P}\left[\overline{X} > (1+\epsilon)\cdot p\right] &\leq \exp\left(\frac{-N\epsilon^2 p}{3}\right) \\
\mathbb{P}\left[\overline{X} < (1-\epsilon)\cdot p\right] &\leq \exp\left(\frac{-N\epsilon^2 p}{2}\right) \\
\implies \mathbb{P}\left[\left|\overline{X} - \mathbb{E}[X]\right| < \epsilon\cdot\mathbb{E}[X]\right] &\leq 2\exp\left(\frac{-N\epsilon^2\cdot\mathbb{E}[X]}{3}\right)
\end{aligned}$$

Proof. We aim to shift the problem to a form so that we can apply Markov's Inequality. Therefore

$$\begin{aligned}
\mathbb{P}\left[\overline{X} < (1-\epsilon)\cdot\mathbb{E}[X]\right] &= \mathbb{P}\left[-s\cdot\overline{X} > -s\cdot(1-\epsilon)\cdot\mathbb{E}[X]\right] && [s > 0 \text{ is a scale parameter}] \\
&= \mathbb{P}\left[\exp\left(-s\cdot\overline{X}\right) > \exp\left(-s\cdot(1-\epsilon)\cdot\mathbb{E}[X]\right)\right] \\
&\leq \frac{\mathbb{E}\left[\exp\left(-s\cdot\overline{X}\right)\right]}{\exp\left(-s\cdot(1-\epsilon)\cdot\mathbb{E}[X]\right)} && [\text{Using Markovs' Inequality}] \\
&= \frac{\prod_{n\in[N]}\mathbb{E}\left[\exp\left(-\frac{s}{n}\cdot X_n\right)\right]}{\exp\left(-s\cdot(1-\epsilon)\cdot\mathbb{E}[X]\right)} \\
&= \left[\frac{1+p(e^{-t}-1)}{\exp\left(-t\cdot(1-\epsilon)p\right)}\right]^N && \left[t = \frac{s}{N} \text{ and } \mathbb{E}[X] = p\right] \\
&\leq \exp\left(Np(e^{-t}-1) + Npt\cdot(1-\epsilon)\right)
\end{aligned}$$

Since we need to find the closest bound, we set the differential with respect to t as 0. Hence $t = -\ln(1-\epsilon)$. Therefore

$$\begin{aligned}
\mathbb{P}\left[\overline{X} < (1-\epsilon)\mathbb{E}[X]\right] &\leq \exp\left[-Np\left(\epsilon + \ln(1-\epsilon)(1-\epsilon)\right)\right] \\
&= \mathrm{ex}\left[-Np\left(\frac{\epsilon^2}{2} - (\dots)\right)\right] && [\text{Using Taylor's Expansion for } \ln(1+x)] \\
&\leq \exp\left(\frac{-N\epsilon^2\cdot\mathbb{E}[X]}{2}\right)
\end{aligned}$$

□

4. **Hoeffding's Inequality**

Exercise 6.2. For the Chernoff's Bound discussed earlier,

(i) prove the second inequality in *i.e.* $\mathbb{P}\left[\overline{X} > (1+\epsilon)\mathbb{E}[X]\right] \leq \exp\left(\frac{-n\epsilon^2\mathbb{E}[X]}{3}\right)$

(ii) find out the term for the Chernoff's Bound if $\{X_n\}$ are independent Rademacher Variables *i.e.* $X_n \in \{-1,1\}$

We can also find a lower bound to the Chernoff's inequalities (Anti-Concentration Bound)

Theorem 6.2 (Lower Bounds for Sampling Algorithms — Canetti, Goldriech)**.** For the same set of random variables as in Chernoff's Bound, we can show that

$$\mathbb{P}\left[\left|\overline{X} - \mathbb{E}[X]\right|\right] \quad \geq \quad \frac{1}{8e\pi}\exp\left(-4\epsilon^2 n\mathbb{E}[X]^2\right)$$