Lecture 10

Instructor: Quan Zhou

For more details about the materials covered in this note, see Chapters 5.2, 6.5 and 10.3 of Resnick [3] and Chapters 1.5, 1.6 and 4.1 of Durrett [2].

10.1 Moments and norms of random variables

Definition 10.1. Variance and covariance of random variables:

$$\mathrm{Var}(X) = E[(X - E[X])^2],$$
$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])],$$

provided that $E[X^2] < \infty$ and $E[Y^2] < \infty$.

Proposition 10.1. *Properties of variance and covariance (assuming $E[X^2] < \infty$ and $E[Y^2] < \infty$).*

(i) $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ *and* $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

(ii) For any $a \in \mathbb{R}$, $\mathrm{Var}(X + a) = \mathrm{Var}(X)$ *and* $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$.

(iii) $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.

(iv) If X, Y are independent, $\mathrm{Cov}(X, Y) = 0$.

Proof. Try it yourself. □

Example 10.1. Let $X \sim N(0, 1)$ be a standard normal random variable. Let Z be independent of X and $\mathsf{P}(Z = 1) = \mathsf{P}(Z = -1) = 0.5$. Define $Y = XZ$. Note that Y is also a standard normal random variable. Further, $\mathrm{Cov}(X, Y) = E[XY] = E[ZX^2] = E[Z] = 0$. But X, Y are not independent.

Definition 10.2. The L^p norm of a random variable is defined by

$$\|X\|_p = (E[|X|^p])^{1/p}, \qquad 1 \leq p < \infty.$$

This notation is also defined for $0 < p < 1$, though $\|\cdot\|_p$ is not a norm for $0 < p < 1$.

Definition 10.3. L^∞ norm: $\|X\|_\infty = \inf\{a \in \mathbb{R} : \mathsf{P}(|X| > a) = 0\}$. If there is no $a \in \mathbb{R}$ such that $\mathsf{P}(|X| > a) = 0$, then $\|X\|_\infty = \infty$. This is also called the essential supremum of $|X|$.

Definition 10.4. Moment generating function of the random variable X:

$$M_X(t) = E[e^{tX}], \qquad t \in \mathbb{R}.$$

Proposition 10.2. *Properties of MGF.*

(i) For independent random variables $X_1, \dots, X_n$ *and constants* $a_1, \dots, a_n$*. The MGF of the sum* $S_n = a_1X_1 + \cdots + a_nX_n$ *is given by*

$$M_{S_n}(t) = M_{X_1}(a_1t)M_{X_2}(a_2t)\cdots M_{X_n}(a_nt).$$

(ii) If $M_X(t)$ *is finite in a neighborhood of* 0*, then*

$$E[X^n] = \frac{d^n M_X(t)}{dt^n}\Big|_{t=0}.$$

(iii) If in a neighborhood of 0*, both* $M_X(t)$ *and* $M_Y(t)$ *are finite and they are equal, then* X, Y *have the same distribution.*

Proof. Part (i) is easy to show. We do not prove the other two parts here. □

10.2 Inequalities involving expectations and moments

Theorem 10.1 (Jensen's inequality)**.** *For any convex function* φ*,* $E[\varphi(X)] \geq \varphi(E[X])$*, provided that both expectations exist. More generally, if* $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ *is convex, then* $E[\varphi(X_1, \dots, X_n)] \geq \varphi(E[X_1], \dots, E[X_n])$ *provided that all the expectations involved exist.*

Proof. See the textbook. □

Example 10.2. Typical examples of convex functions include: $|x|$, x^2, e^x, $|x|^p$ for $p \geq 1$. Typical examples of concave functions include: $\sqrt{x}$, $\log x$. Affine functions of the form $ax + b$ are both convex and concave. Hence, by Jensen's inequality, $E[X^2] \geq (E[X])^2$ (i.e. $\mathrm{Var}(X) \geq 0$) and $E|X| \geq |EX|$.

Example 10.3. Consider the function $f(x) = |x|^{q/p}$ where $0 < p \leq q < \infty$. Then f is convex. For a random variable X with finite absolute moments of any order, define another random variable $Y = |X|^p$. By Jensen's inequality,

$$E(|Y|^{q/p}) \geq |EY|^{q/p} \quad \Longrightarrow \quad \left\{E(|Y|^{q/p})\right\}^{1/q} \geq |EY|^{1/p}.$$

Substituting $Y = |X|^p$, we get $\|X\|_p \leq \|X\|_q$.

Theorem 10.2 (Hölder's inequality)**.** *For $1 \leq p, q \leq \infty$ s.t. $1/p + 1/q = 1$,*

$$E|XY| \leq \|X\|_p \|Y\|_q.$$

When $p = q = 2$, this is known as Cauchy-Schwarz inequality.

Proof. See the textbook. □

Example 10.4. By Cauchy-Schwarz inequality, $|\mathrm{Cov}(X, Y)|^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$.

Theorem 10.3 (Minkowski inequality)**.** *For any $1 \leq p \leq \infty$ and X, Y s.t. $\|X\|_p < \infty, \|Y\|_p < \infty$, we have the "triangle inequality",*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Proof. Try it yourself. □

Theorem 10.4 (Markov inequality)**.** *For a non-negative random variable X and any $a > 0$, $\mathsf{P}(X \geq a) \leq E[X]/a$.*

Proof. Note that $\mathsf{P}(X \geq a) = E[\mathbb{1}_{\{X \geq a\}}]$, and on the set $\{X \geq a\}$, we have $X/a \geq 1$ since $a > 0$. Thus,

$$E[\mathbb{1}_{\{X \geq a\}}] \leq E\left[\frac{X}{a}\mathbb{1}_{\{X \geq a\}}\right] = \frac{E[X\mathbb{1}_{\{X \geq a\}}]}{a} \leq \frac{E[X]}{a},$$

which completes the proof. □

Corollary 10.1 (Chebyshev inequality)**.** $\mathsf{P}(|X - E(X)| \geq a) \leq \mathrm{Var}(X)/a^2$ *for any $a > 0$.*

Proof. Try it yourself. □

Corollary 10.2 (Chernoff bound)**.** $\mathsf{P}(X \geq a) \leq e^{-ta}E[e^{tX}]$ *for any $t > 0$.*

Proof. Try it yourself. □

Example 10.5. Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. random variables with $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Clearly, $E(\bar{X}_n) = \mu$. The variance of $\bar{X}_n$ can be computed by

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{\sigma^2}{n},$$

where we have used the assumption that X_i's are i.i.d. By Chebyshev inequality, for any $c > 0$,

$$\mathsf{P}(|\bar{X}_n - \mu| \geq c) \leq \frac{\mathrm{Var}(\bar{X}_n)}{c^2} = \frac{\sigma^2}{nc^2} \to 0,$$

as $n \to \infty$. This is a special case of Weak Law of Large Numbers.

Example 10.6. Let $X \sim N(0, 1)$. One can show that the MGF is given by $E[e^{tX}] = e^{t^2/2}$. Hence, use the Chernoff bound with $t = x$, we get

$$\mathsf{P}(X > x) \leq e^{-tx} E[e^{tX}] = \exp\left(\frac{t^2 - 2tx}{2}\right) = e^{-x^2/2}.$$

10.3 Inequalities involving conditional expectations

For all results below, assume X is an *integrable* random variable defined on $(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{G} \subset \mathcal{F}$ is a given sub-σ-algebra.

Proposition 10.3 (Jensen's inequality for conditional expectation)**.** *If φ is convex and $E|\varphi(X)| < \infty$, then*

$$\varphi(E[X \mid \mathcal{G}]) \leq E[\varphi(X) \mid \mathcal{G}], \quad a.s.$$

Proof. See the textbook. □

Theorem 10.5 (Conditional expectation as a projection)**.** *Let $L^2(\Omega, \mathcal{G}, \mathsf{P}) = \{Y : Y \in \mathcal{G},\ EY^2 < \infty\}$. If $EX^2 < \infty$, then $\inf_{Y \in L^2(\Omega, \mathcal{G}, \mathsf{P})} E(X - Y)^2$ is attained by $Y = E[X \mid \mathcal{G}]$.*

Proof. Try it yourself. □

Theorem 10.6 (Conditional expectation as a contraction)**.** *Suppose $p \geq 1$. If $E|X|^p < \infty$, then $\|E[X \mid \mathcal{G}]\|_p \leq \|X\|_p$.*

Proof. Try it yourself. □

10.4 Union bound

The following result is often known as union bound or Boole's inequality: for a countable sequence of events $A_1, A_2, \ldots$, $\mathsf{P}(\cup_i A_i) \leq \sum_i \mathsf{P}(A_i)$. But it is just a paraphrase of the σ-subadditivity of measures.

Example 10.7. If $\mathsf{P}(A_i) = \alpha/n$ for $i = 1, 2, \ldots, n$. Then $\mathsf{P}(\cup A_i) \leq \alpha$ by the union bound. This is why we do Bonferroni correction in multiple testing.

Example 10.8. By De Morgan's laws and union bound,

$$\mathsf{P}(\cap_i A_i) = 1 - \mathsf{P}(\cup A_i^c) \geq 1 - \sum_{i=1}^{n} (1 - \mathsf{P}(A_i)) = \sum_{i=1}^{n} \mathsf{P}(A_i) - (n-1).$$

10.5 Tail bound for normal distribution

We have already obtained a tail bound for normal random variables in Example 10.6. But very often we can find tail bounds for random variables using more elementary techniques: differentiation and integration by parts.

Theorem 10.7. *Let $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ be the density function of the standard normal distribution. Then, for $x > 0$,*

$$\frac{x}{x^2+1}\phi(x) \leq \int_x^\infty \phi(t)dt \leq \frac{1}{x}\phi(x).$$

Proof. The lower bound can be proven by differentiating (try it yourself). To prove the upper bound, use integration by parts to get

$$\begin{aligned} \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt &= \int_x^\infty -\frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} de^{-t^2/2} \\ &= -\frac{\phi(t)}{t}\Big|_x^\infty - \int_x^\infty \frac{1}{t^2}\phi(t)dt \\ &= \frac{\phi(x)}{x} - \int_x^\infty \frac{1}{t^2}\phi(t)dt \leq \frac{\phi(x)}{x}. \end{aligned}$$

Note that the trick of integration by parts can be applied again to bound the remainder term, $\int_x^\infty (\phi(t)/t^2)dt$. □

References

[1] Dennis D. Cox. *The Theory of Statistics and Its Applications*. Unpublished.

[2] Rick Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 2019.

[3] Sidney Resnick. *A Probability Path*. Springer, 2019.