STATS 310A: Theory of Probability I **Autumn 2016/17**

Lecture 12: November 2

Lecturer: Persi Diaconis *Scribes: Kenneth Tay*

12.1 Expectations

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition 12.1 *A* ***random variable*** *X is a measurable function $X : \Omega \to (\mathbb{R}, \text{Borel sets})$, and, if the integral exists, the* ***expectation*** *of X is*

$$\mathbb{E}(X) := \int_\Omega X(\omega) P(d\omega) = \int_{-\infty}^{\infty} xF(dx),$$

where $F = P^{X^{-1}}$.

(Note that if $X(\omega) = \delta_A(\omega)$, then $\mathbb{E}(X) = P(A)$.)

Main problem of probability: Given $(\Omega, \mathcal{F}, P)$ and a random variable X, compute or approximate $\mathbb{E}X$.

12.1.1 Sum of independent random variables

Let X and Y be independent random variables. Find the distribution of $X + Y$.

Define 2 probability measures on $\mathbb{R}$: $\mu(A) = P(X \in A)$ and $\nu(B) = P(Y \in B)$. By independence, we have

$$\mu \times \nu(C) = \int_{-\infty}^{\infty} \mu(C_y)\nu(dy)$$

for any set C in the product space.

Let $D \subseteq \mathbb{R}$, and let $C := \{(x, y) : x + y \in D\} = D - y = D - x$. Then

$$P\{X + Y \in D\} = \int_{-\infty}^{\infty} \mu(D - y)\nu(dy) =: \mu * \nu(D).$$

Examples:

1. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\eta)$ be independent. Then

$$P\{X \in A\} = \sum_{j \in A} \frac{e^{-\lambda}\lambda^j}{j!}, \quad P\{X \in B\} = \sum_{j \in B} \frac{e^{-\eta}\eta^j}{j!}.$$

If $D = \{l\}$, then

$$\begin{aligned} P\{X+Y=l\} &= \sum_{a=0}^{l} \frac{e^{-\lambda}\lambda^a}{a!} \frac{e^{-\eta}\eta^{l-a}}{(l-a)!} \\ &= \frac{e^{-\lambda-\eta}\eta^l}{l!} \sum_{a=0}^{l} \left(\frac{\lambda}{\eta}\right)^a \binom{l}{a} \\ &= \frac{e^{-\lambda-\eta}\eta^l}{l!} \left(1+\frac{\lambda}{\eta}\right)^l \\ &= \frac{e^{-\lambda-\eta}(\lambda+\eta)^l}{l!}. \end{aligned}$$

So $X+Y \sim \text{Poisson}(\lambda+\eta)$.

2. If $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y \sim \mathcal{N}(\nu, \tau^2)$, independent, then $X+Y \sim \mathcal{N}(\mu+\nu, \sigma^2+\tau^2)$.
3. Let $X \sim \Gamma(a)$ and $Y \sim \Gamma(b)$ be independent, i.e.

$$P(X \in A) = \int_A \frac{x^{a-1}e^{-x}}{\Gamma(a)} dx.$$

Then $\dfrac{X}{X+Y}$ is independent of $X+Y$, and $X+Y \sim \Gamma(a+b)$, $\dfrac{X}{X+Y} \sim \text{Beta}(a,b)$, i.e.

$$P\left(\frac{X}{X+Y} \in A\right) = \int_A \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} dx.$$

4. If X, Y, Z are independent with distributions $\Gamma(a), \Gamma(b), \Gamma(c)$ respectively, then $\dfrac{X}{X+Y+Z}, \dfrac{X+Y}{X+Y+Z}$ and $X+Y+Z$ are independent with distributions $\text{Beta}(a, b+c)$, $\text{Beta}(a+b, c)$, and $\Gamma(a+b+c)$ respectively. (For details, see Hogg & Craig, "Introduction to Probability".)

The proposition below gives a useful formula for calculating expectations:

Proposition 12.2 *Suppose $X \geq 0$. Then*

$$\mathbb{E}X = \int_0^\infty P(X \geq t)dt = \int_0^\infty P(X > t)dt.$$

Proof:

Step 1: Assume X is simple, i.e. taking on a finite number of values $0 \leq x_1 < x_2 < \cdots < x_k$.

$$\begin{aligned} \mathbb{E}X &= \sum_{j=1}^{k} x_j P\{X = x_j\} \\ &= \sum_{j=1}^{k-1} x_j [P\{X \geq x_j\} - P\{X \geq x_{j-1}\}] + x_k P(X \geq x_k) \\ &= x_1 P(X \geq x_1) + \sum_{j=2}^{k} (x_j - x_{j-1}) P(X \geq x_j). \end{aligned}$$

Note that

$$P(X \geq x) = \begin{cases} P(X \geq x_1) & \text{for } 0 < x \leq x_1, \\ P(X \geq x_j) & \text{for } x_{j-1} < x \leq x_j. \end{cases}$$

Hence, thinking of the last sum above as a Riemann sum, we have

$$\mathbb{E}X = x_1 P(X \geq x_1) + \sum_{j=2}^{k} (x_j - x_{j-1}) P(X \geq x_j) = \int_0^\infty P(X \geq x)dx.$$

Step 2: For $X \geq 0$, find a sequence of simple functions X_n such that $X_n \nearrow X$ pointwise. By the Monotone Convergence Theorem,

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}(\lim X_n) \\ &= \lim \mathbb{E}X_n \\ &= \lim \int_0^\infty P(X_n > t)dt \\ &= \int_0^\infty \lim P(X_n > t)dt \\ &= \int_0^\infty P(X > t)dt. \end{aligned}$$

■

Example: Guessing cards. (This example has roots in card guessing to show whether someone has ESP or not.)

Say we have n cards labeled $1, 2, \ldots, n$. I shuffle the cards and you guess the cards one by one. We want to compute the expected number of right guesses under different circumstances.

- **Experiment 1:** No feedback (you are not told whether your guess was right or wrong).

  For any card, the chance of getting it right is $1/n$, so the expected number of right guesses is $\frac{1}{n} \cdot n = 1$.

- **Experiment 2:** Complete feedback (after each guess, you are told what the card actually is).

  The chance of getting the first card right is $1/n$, the chance of getting the second right is $1/(n-1)$, etc. Thus, the expected number of right guesses is

  $$\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1} = H_n \approx \log n.$$

  (For $n = 52$, $H_n \approx 4.5$.)

- **Experiment 3:** Yes/no feedback (after each guess, you are told whether you guessed correctly or not). It turns out that the optimal strategy is the following: Guess a card (say 1), and keep guessing it until you get it right. Then guess another card (say 2), and keep guessing, and so on.

  Let X be the number of correct guesses under the optimal strategy. Then $P(X \geq 1) = 1$, $P(X \geq 2) = \frac{1}{2}$, $P(X \geq 3) = \frac{1}{6}$. In general, if the cards I guess are $1, 2, \ldots,$ $X \geq k$ iff the cards $1, 2, \ldots, k$ appear in that order in the deck. So $P(X \geq k) = \frac{1}{k!}$. Thus,

  $$\mathbb{E}X = \sum_{k=1}^{\infty} \frac{1}{n!} = e - 1 + O\left(\frac{1}{n!}\right) \approx 1.7.$$

Theorem 12.3 *Let X and Y be independent random variables with finite expectations. Then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Proof:

Step 1: X and Y are indicator functions.

Let $X = \delta_A$, $Y = \delta_B$. Then $XY = \delta_{A\cap B}$, and by definition, $P(A \cap B) = P(A)P(B)$.

Step 2: Simple functions. If $X = \sum_{i=1}^{n} a_i \delta_{A_i}$, $Y = \sum_{j=1}^{m} b_j \delta_{B_j}$, then

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{E}(\delta_{A_i}\delta_{B_j}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{E}(\delta_{A_i})\mathbb{E}(\delta_{B_j}) \\
&= \sum_{i=1}^{n} a_i \mathbb{E}(\delta_{A_i}) \sum_{j=1}^{m} b_j \mathbb{E}(\delta_{B_j}) \\
&= \mathbb{E}X\mathbb{E}Y.
\end{aligned}$$

Step 3: $X, Y \geq 0$.

Pick 2 sequences of simple functions $\{X_n\}$ and $\{Y_n\}$ such that $X_n \nearrow X$ and $Y_n \nearrow Y$. By Step 2 and the Monotone Convergence Theorem, we are done.

Step 4: General X, Y.

Write $X = X^+ - X^-$, $Y = Y^+ - Y^-$. Note that X^+ and X^- are independent of Y^+ and Y^-.

Since the expectations of these 4 new functions are finite, and $XY = X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-$, we can take expectations and use Step 3 to get the result. ∎

12.2 Moment Generating Functions

Definition 12.4 *If X is a random variable, then the* ***k^{th} moment****, if it exists, is*

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k F(dx).$$

Example: If $X \sim \mathcal{N}(0,1)$, then

$$\begin{aligned}
\mathbb{E}(X^k) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^k e^{-x^2/2} dx \\
&= \begin{cases} 0 & \text{if } k \text{ odd}, \\ \dfrac{(2j)!}{2^j j!} & \text{if } k = 2j. \end{cases}
\end{aligned}$$

Definition 12.5 *Let X be a random variable, then the* ***moment generating function*** *of X is*

$$M(s) = \mathbb{E}\left(e^{sX}\right)$$

for $s \in \mathbb{R}$, if the expectation exists.

Note that we always have $M(0) = 1$, and if $X \geq 0$, then $M(s) < \infty$ for $s \leq 0$.

Example: $X \sim \text{Poisson}(\lambda)$. Then, for all s,

$$\begin{aligned}
\mathbb{E}e^{sX} &= \sum_{j=0}^{\infty} e^{sj} \frac{e^{-\lambda}\lambda^j}{j!} \\
&= e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda e^s)^j}{j!} \\
&= e^{-\lambda} e^{\lambda e^s} \\
&= \exp[\lambda(e^s - 1)].
\end{aligned}$$

Suppose $M(s)$ is finite on some interval $(-s_0, s_0)$ where $s_0 > 0$. Then, since $e^{|sx|} \leq e^{sx} + e^{-sx}$ with both functions on the RHS integrable, we have

$$\sum_{k=0}^{\infty} \frac{|sx|^k}{k!} = e^{|sx|}$$

for $s \in (-s_0, s_0)$. Taking expected values on both sides,

$$\sum_{k=0}^{\infty} s^k \frac{\mathbb{E}|X^k|}{k!} = E\left(e^{|sX|}\right).$$

Since the RHS is finite, we conclude that all the moments are finite and

$$\sum_{k=0}^{\infty} s^k \frac{\mathbb{E}X^k}{k!} = E\left(e^{sX}\right).$$

Using the Dominated Convergence Theorem, we can obtain the following proposition which allows us to compute moments from the m.g.f.:

Proposition 12.6 *$M(s)$ is infinitely differentiable and $M^{(k)}(0) = \mathbb{E}(X^k)$.*

The following proposition helps us to compute the m.g.f. for sums of independent random variables:

Proposition 12.7 *If $X_1, \ldots, X_n$ are independent random variables with m.g.f.s $M_1, \ldots, M_n$ which are finite at s, then*

$$M_{X_1+\cdots+X_n}(s) = \prod_{i=1}^{n} M_i(s).$$

Consider the set of all positive random variables. This set is a semi-group under convolution with identity $= \delta_{\{0\}}$. In this context, $M(s) = \mathbb{E}e^{-sX}$ is a homomorphism onto "complete monotone functions" (i.e. $(-1)^k M^{(k)}(s) \geq 0$ for all k, $s \in (0, \infty)$, and $M(0) = 1$).