STATS310A - Lecture 11

Persi Diaconis
Scribed by Michael Howes

10/26/21

Contents

1 Expectation

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose $X : \Omega \to \mathbb{R}$ is measurable. We define *the expectation of X* to be

$$\mathbb{E}[X] = \int X(\omega)\mathbb{P}(d\omega),$$

whenever the above intergal exists.

Note that if $X = \delta_A$, then $\mathbb{E}[X] = \mathbb{P}(A)$. Thus we can once again reframe our fundamental question of probability:

Given a random variable X, compute or approximate $\mathbb{E}[X]$.

which generalizes our previous versions of this question.

1.1 Sums of random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose X, Y are independent random variables on Ω. Define the following probabilities on $\mathbb{R}$,

$$\mu(A) = \mathbb{P}(X \in A) \text{ and } \nu(A) = \mathbb{P}(Y \in A).$$

Since X, Y are independent, we know that for all measurable $C \subseteq \mathbb{R}^2$,

$$\mathbb{P}((X, Y) \in C) = (\mu \times \nu)(C),$$

where $\mu \times \nu$ is the product measure given by

$$(\mu \times \nu)(C) = \int_{\mathbb{R}} \mu(C_y)\nu(dy) = \int_{\mathbb{R}} \nu(C_x)\mu(dx).$$

Given $D \subseteq \mathbb{R}$ measurable, define $C = \{(x,y) : x + y \in D\} \subseteq \mathbb{R}^2$. Note that $C_y = D - y$ and $C_x = D - x$. Thus

$$\begin{aligned}
\mathbb{P}(X + Y \in D) &= \mathbb{P}((X,Y) \in C) \\
&= \int_{\mathbb{R}} \nu(C_x)\mu(dx) \\
&= \int_{\mathbb{R}} \nu(D - x)\mu(dx).
\end{aligned}$$

And likewise

$$\mathbb{P}(X + Y \in D) = \int_{\mathbb{R}} \mu(D - y)\nu(dy).$$

The above defines a measure on $\mathbb{R}$ and is called the convolution of μ and ν. It is denoted $\mu * \nu$.

Example 1. Suppose $\mu = \text{Poisson}(\theta)$ and $\nu = \text{Posson}(\eta)$ for some $\theta, \eta \geq 0$. That is

$$\mu(A) = \sum_{j \in A} \frac{e^{-\theta}\theta^j}{j!} \text{ and } \nu(B) = \sum_{j \in B} \frac{e^{-\eta}\eta^j}{j!},$$

where $A, B \subseteq \{0, 1, 2, \ldots\}$. Note that for $l = 0, 1, 2, \ldots$,

$$\begin{aligned}
(\mu * \nu)(\{l\}) &= \sum_{a=0}^{l} \frac{e^{-\theta}\theta^a}{a!} \frac{e^{-\eta}\eta^{l-a}}{(l-a)!} \\
&= \frac{e^{-(\theta+\eta)}\eta^l}{l!} \sum_{a=0}^{l} \binom{l}{a} \left(\frac{\theta}{\eta}\right)^a \\
&= \frac{e^{-(\theta+\eta)}\eta^l}{l!} \left(1 + \frac{\theta}{\eta}\right)^l \\
&= \frac{e^{-(\theta+\eta)}(\theta+\eta)^l}{l!}.
\end{aligned}$$

Thus $\mu * \nu = \text{Poisson}(\theta + \eta)$.

Example 2. Similarly if $X \sim \mathcal{N}(\theta_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\theta_2, \sigma_2^2)$ and X, Y are independent, then $X + Y \sim \mathcal{N}(\theta_1 + \theta_2, \sigma_1^2 + \sigma_2^2)$.

2 Homework

Read chapters 21-22 and do problem 20.8, 20.9, 21.3 and 21.6. There are two other problems which we describe now.

2.1 Problem (A)

This problem is about the $\beta - \Gamma$ calculus. For $\alpha > 0$, we say $X \sim \text{Gamma}(\alpha)$ on $[0, \infty)$ if X has density

$$\frac{e^{-x}x^{\alpha-1}}{\Gamma(\alpha)}, \text{ for } x \geq 0.$$

For $\alpha, \beta > 0$, we say that $W \sim \text{Beta}(\alpha, \beta)$ on $[0, 1]$ is W has density

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ for } 0 \leq x \leq 1.$$

On the homework we will show:

(a) If $X \sim \text{Gamma}(\alpha)$ and $Y \sim \text{Gamma}(\beta)$ and X is independent of Y, then $\frac{X}{X+Y}$ and $X + Y$ are independent and $\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$, $X + Y \sim \text{Gamma}(\alpha + \beta)$.

(b) If $X, Y, Z \sim \text{Gamma}(\alpha), \text{Gamma}(\beta), \text{Gamma}(\gamma)$ are independent, then

$$\frac{X}{X+Y}, \frac{X+Y}{X+Y+Z} \text{ and } X + Y + Z,$$

are independent and they have distributions $\text{Beta}(\alpha, \beta)$, $\text{Beta}(\alpha + \beta, \gamma)$ and $\text{Gamma}(\alpha + \beta + \gamma)$ respectively.

2.2   Problem (B)

Before discussing problem (B) on the homework. We will state, prove and apply a theorem.

Theorem 1. *Suppose $X \geq 0$, then*

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

Note that the two integrals are indeed equal since $\mathbb{P}(X = t) > 0$ for at most countably many t. Thus the functions $\mathbb{P}(X \geq t)$ and $\mathbb{P}(X > t)$ agree almost everywhere with respect to Lebesgue measure.

Proof. Suppose that $X = \sum_{j=1}^k x_j \delta_{A_j}$ with $0 \leq x_1 \leq x_2 \leq \ldots \leq x_k$. Then

$$\mathbb{E}[X] = \sum_{j=1}^k x_j \mathbb{P}(X = x_j)$$

$$= \sum_{j=1}^{k-1} x_j \left(\mathbb{P}(X \geq x_j) - \mathbb{P}(X > x_{j+1})\right) + x_k \mathbb{P}(X \geq x_k)$$

$$= x_1 \mathbb{P}(X \geq x_1) + \sum_{j=1}^k (x_j - x_{j-1})\mathbb{P}(X \geq x_j).$$

Observe $\mathbb{P}(X \geq t) = \mathbb{P}(X \geq x_1) = 1$ for $t \in [0, x_1]$ and $\mathbb{P}(X \geq t) = 0$ for $t > x_k$. Also note that if $x_{j-1} < t \leq x_j$, then

$$\mathbb{P}(X \geq t) = \mathbb{P}(X \geq x_j).$$

Thus the function $t \mapsto \mathbb{P}(X \geq t)$ is a step function and we have

$$\int_0^\infty \mathbb{P}(X \geq t)dt = x_1 + \mathbb{P}(X \geq x_i) + \sum_{j=1}^k (x_j - x_{j-1})\mathbb{P}(X \geq x_j) = \mathbb{E}[X].$$

A limiting argument via the monotone convergence theorem allows us to conclude that the result holds for all $X \geq 0$. $\square$

2.2.1   ESP

We can apply this theorem in an interesting setting. Suppose that someone claims to have ESP (extrasensory preception). We could test their claim by asking them what is on a card without letting them see the card. They may get some cards right just by chance. How many cards they get right by chance will depend on how much feedback they are given. Suppose that there are n cards and they are all distinct. Suppose that the deck is well shuffled and we ask them to guess each card. The three types of feedback we can give are

- No feedback.

- Complete feedback. After they guess we tell them if they were right or wrong and which card was drawn.

- Yes/No feedback. We tell them if their guess is correct or incorrect but we do not reveal cards if they are incorrect.

We can ask what is the expected number of correct guesses for each type of feedback assuming that the person does not have ESP. Let $X_i = 1$ if the person correctly guesses card i and let $X_i = 0$ otherwise. Define

$$S_n = \sum_{i=1}^{n} X_i.$$

We wish to calculate $\mathbb{E}[S_n]$. For no feedback they have a $\frac{1}{n}$ chance of correctly guessing the ith card. Thus the expected number of correct guesses is

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \frac{1}{n} = \frac{n}{n} = 1.$$

For complete feedback, the person has a $\frac{1}{n}$ chance of guessing the first card, a $\frac{1}{n-1}$ chance of guessing the second, a $\frac{1}{n-2}$ chance of guessing the third and so on. Thus

$$\begin{aligned}
\mathbb{E}[S_n] &= \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{2} + 1 \\
&= H(n) \\
&= \log(n) + \gamma + O(1/n).
\end{aligned}$$

In particular when $n = 52$, $\mathbb{E}[S_n] \approx 4$. In the case of yes/no feedback one can show that the optimal (ESP-free) stratergy is

- Guess card 1 until correct.

- Guess card 2 until correct.

- Guess card 3 . . . .

This claim requires proof and we won't prove it here but it is definitely believable. Under this stratergy we can use the previous theorem to calculate $\mathbb{E}[S_n]$. Note that $\mathbb{P}(S_n \geq 1) = 1$, since we will always get the first card right. Also $S_n \geq 2$ if and only if card 1 is above card 2 and so $\mathbb{P}(S_n \geq 2) = \frac{1}{2}$. In general $S_n \geq k$ if and only if cards $1, 2, \ldots, k$ are in increasing order. This occurs with probability $\frac{1}{k!}$.

Thus, by the theorem

$$
\begin{aligned}
\mathbb{E}[S_n] &= \sum_{k=1}^{n} \mathbb{P}(S_n \geq k) \\
&= \sum_{k=1}^{n} \frac{1}{k!} \\
&= e - 1 + O(1/n!) \\
&\approx 1.7.
\end{aligned}
$$

2.2.2 The homework problem

Let $H(x, y)$ be the bivariate distribution function for a random vector (X, Y). Suppose H has marginals

$$H(x, \infty) = F(x) \text{ and } H(\infty, y) = G(y).$$

Suppose also that $X, Y \geq 0$. On the homework we must prove

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \int_0^\infty \int_0^\infty \mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)\mathbb{P}(Y \geq t) ds dt.$$

Recall the bivariate distributions

$$L(x, y) = (F(x) + G(y) - 1)_+$$

and

$$U(x, y) = F(x) \wedge G(y).$$

On the homework we must also show that L is the bivariate distribution that minimizes the covariance between X and Y and U is the bivariate distribution that maximizes the covariance between X and Y. The distribution L and U are called the Frechet's extremal distributions of F and G.

3 Independence

Theorem 2. *If X and Y are independent and integrable random varaibles, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Proof. We will proceed in stages (0), (1), (2), (3) (and break slightly from our usual order).

Step (0): We will prove the result if $X = \delta_A$ and $Y = \delta_B$. Note that in this case $XY = \delta_{A \cap B}$ and thus

$$\mathbb{E}[XY] = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}[X]\mathbb{E}[Y],$$

by independence.

Step (1): Suppose $Y = \sum_{i=1}^{m} a_i \delta_{A_i}$ and $X = \sum_{j=1}^{n} b_j \delta_{B_j}$. Then

$$XY = \sum_{i=1}^{m} \sum_{j=1}^{n} \delta_{A_i} \delta_{B_j}.$$

By linearity and step 0 we have

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

Step (2): suppose that $X, Y \geq 0$, then there exist $X_n, Y_n \geq 0$ which are independent and of the form of step (1) and $X_n \nearrow X$, $Y_n \nearrow Y$. Thus $X_n Y_n \nearrow XY$ and by the monotone convergence theorem

$$
\begin{aligned}
\mathbb{E}[XY] &= \lim_{n} \mathbb{E}[X_n Y_n] \\
&= \lim_{n} \mathbb{E}[X_n]\mathbb{E}[Y_n] \\
&= \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}
$$

Step (3): consider general X and Y and write $X = X + -X_-$ and $Y = Y_+ - Y_-$. Then

$$
\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[X_+ Y_+ - X_+ Y_- - X_- Y_+ + X_- Y_-] \\
&= \mathbb{E}[X_+]\mathbb{E}[Y_+] - \mathbb{E}[X_+]\mathbb{E}[Y_-] - \mathbb{E}[X_-]\mathbb{E}[Y_+] + \mathbb{E}[X_-]\mathbb{E}[Y_-] \\
&= \mathbb{E}[X_+ - X_-]\mathbb{E}[Y_+] - \mathbb{E}[X_+ - X_-]\mathbb{E}[Y_-] \\
&= \mathbb{E}[XY]. \qquad \qquad \square
\end{aligned}
$$

4 Moments

Definition 2. A random variable X has an ith *moment* μ_i if $\mathbb{E}[|X|^i] < \infty$ and $\mathbb{E}[X^i] = \mu_i$.

Example 3. If $X \sim \mathcal{N}(0, 1)$, then

$$
\mathbb{E}[X^i] = \int_{-\infty}^{\infty} \frac{x^i}{\sqrt{2\pi}} e^{-\frac{1}{2} x^i} dx = \begin{cases} 0 & \text{if } i \text{ is odd,} \\ (i-1)!! & \text{if } i \text{ is even.} \end{cases}
$$

The notation $(i - 1)!!$ is the *skip factorial*. It is recursively defined $n!! = n(n - 2)!!$ for $n \geq 2$ and $0!! = 1!! = 1$.

Definition 3. Given a random variable X with distribution μ, the *moment generating function* of X is defined to be

$$
M_X(s) = \mathbb{E}[e^{sX}] = \int e^{sX} \mu(dx),
$$

provided the above integral exists.

Note $M_X(0) = \mathbb{E}[1] = 1$ for all X and that if $X \geq 0$, the $M_X(s) \leq M_X(r)$ for all $s \leq r$.

Example 4. Suppose $X \sim \text{Poisson}(\lambda)$, then

$$
\begin{aligned}
M_X(s) &= \sum_{j=0}^{\infty} \frac{e^{sj} e^{-\lambda} \lambda^j}{j!} \\
&= e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda e^s)^j}{j!} \\
&= e^{-\lambda} e^{e^s \lambda} \\
&= e^{\lambda(e^s - 1)}.
\end{aligned}
$$

In particular $M_X(s)$ exists for every s.

Suppose that $M(s) < \infty$ for on $[-s_0, s_0]$ for some $s_0 > 0$. Since $e^{|x|} \le e^{-x} + e^x$ for all $x \in \mathbb{R}$ we thus have so $s \in [-s_0, s_0]$,

$$
\mathbb{E}[e^{s|X|}] \le M_X(-s_0) + M_X(s_0).
$$

Thus by the monotone convergence theorem

$$
\mathbb{E}\left[\sum_{j=0}^{\infty} \frac{|sX|^j}{j!} \right] = \sum_{j=0}^{\infty} \frac{|s|^j \mathbb{E}[|X|^j]}{j!}.
$$

And so for $s \in [s_0, s_0]$

$$
M_X(s) = \sum_{j=0}^{\infty} \frac{s^j \mathbb{E}[X^j]}{j!}.
$$

Thus $\mathbb{E}[X^j] = M_X^{(j)}(0)$.

Remark 1. Even though we know the moment generating function for a Poisson random variable, it is still hard to calculate the moments since the derivatives get complicated. We do have that if $X \sim \text{Poisson}(\lambda)$, then

$$
\mathbb{E}[X(X-1)\ldots(X-j+1)] = \lambda^j.
$$

5　A non-Borel set

The following is an example of a subset of $[0,1]$ that is not Borel but is measurable. We will use the following fact. If $x \in [0,1]$ is irrational, then x can be uniquely represented as

$$
x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}},
$$

where $a_i \in \mathbb{N}$. We write this as $x = [a_1, a_2, \ldots]$. Define

$$
A = \{[a_1, a_2, a_3, \ldots] : a_1 \mid a_2 \mid a_3 \mid \ldots\}.
$$

So $[j, j, j, j, \ldots] \in A$ and $[1, 2, 4, 8, 16, \ldots] \in A$ but $[1, 2, 1, 2, \ldots] \notin A$ and no rational numbers are in A. The set A is not Borel but it is Lebesgue measurable. If you are interested in this sort of stuff you should look up the book/topic Descriptive Set Theory.