More on Random Variables and Expectation

RA-MIRI QT 2020-2021

Jensen's inequality

Recall $f : \mathbb{R} \to \mathbb{R}$ is convex if,
for each $x_1, x_2 \in \mathbb{R}$ and for each $t \in [0,1]$ we have

$$f(t\,x_1 + (1-t)x_2) \le t\,f(x_1) + (1-t)\,f(x_2).$$

If f is twice differentiable, a necessary and sufficient condition for f to be convex is that f'' $f''(x) \ge 0$ for $x \in \mathbb{R}$.

Lemma If f is convex then $\mathbf{E}[f(X)] \ge f(\mathbf{E}[X])$.
Proof. Let $\mu = \mathbf{E}[X]$ $(\mu \in \mathbb{R})$. Using Taylor to expand f at $X = \mu$,

$$\begin{aligned} f(X) &= f(\mu) + f'(\mu)(X-\mu) + \frac{f''(\mu)(X-\mu)^2}{2} + \cdots \\ &\ge f(\mu) + f'(\mu)(X-\mu) \\ \mathbf{E}[f(X)] &\ge \mathbf{E}\left[f(\mu) + f'(\mu)(X-\mu)\right] \\ &= \mathbf{E}[f(\mu)] + f'(\mu)(\mathbf{E}[X]-\mu) = f(\mu) \qquad \square \end{aligned}$$

i.e $\mathbf{E}[f(X)] \ge f(\mathbf{E}[X])$.

Expectation of combinations of r.v.

Consider the following experiment:

$X = \text{Unif}(\{1, 2\})$ and $Y = \text{Unif}(\{1, X + 1\})$

(Y depends on X)

What is the expectation of the r.v. XY?

Expectation of combinations of r.v.

Consider the following experiment:

$X = \text{Unif}(\{1, 2\})$ and $Y = \text{Unif}(\{1, X + 1\})$

(Y depends on X)

What is the expectation of the r.v. XY?

$\Omega = \{(1, 1), (1, 2), (2, 1), (2, 2), (2, 3)\}$

$$\mathbf{E}[XY] = \sum_{\omega \in \Omega} X(\omega) Y(\omega) \mathbf{Pr}[\omega]$$

Expectation of combinations of r.v.

Consider the following experiment:

$X = \text{Unif}(\{1, 2\})$ and $Y = \text{Unif}(\{1, X + 1\})$

(Y depends on X)

What is the expectation of the r.v. XY?

$\Omega = \{(1,1), (1,2), (2,1), (2,2), (2,3)\}$

$$\mathbf{E}[XY] = \sum_{\omega \in \Omega} X(\omega) Y(\omega) \mathbf{Pr}[\omega]$$

We have

$$\mathbf{Pr}[(1,1)] = \mathbf{Pr}[(1,2)] = 1/4;$$
$$\mathbf{Pr}[(2,1)] = \mathbf{Pr}[(2,2)] = \mathbf{Pr}[(2,3)] = 1/6.$$

$$\mathbf{E}[XY] = \frac{1}{4} \cdot 1 \cdot 1 + \frac{1}{4} \cdot 1 \cdot 2 + \frac{1}{6} \cdot 2 \cdot 1 + \frac{1}{6} \cdot 2 \cdot 2 + \frac{1}{6} \cdot 2 \cdot 3 = \frac{11}{4}.$$

We have, $\mathbf{Pr}\left[X = 1\right] = 1/2$; $\mathbf{Pr}\left[X = 2\right] = 1/2$ and

$$\mathbf{Pr}\left[Y = 1\right] = \mathbf{Pr}\left[Y = 1 | X = 1\right] + \mathbf{Pr}\left[Y = 1 | X = 2\right] = 1/4 + 1/6 = 5/12$$
$$\mathbf{Pr}\left[Y = 2\right] = \mathbf{Pr}\left[Y = 2 | X = 1\right] + \mathbf{Pr}\left[Y = 2 | X = 2\right] = 1/4 + 1/6 = 5/12$$
$$\mathbf{Pr}\left[Y = 3\right] = \mathbf{Pr}\left[Y = 3 | X = 1\right] + \mathbf{Pr}\left[Y = 3 | X = 2\right] = 0 + 1/6 = 1/6.$$

Then $\mathbf{E}\left[X\right] = 3/2$ and $\mathbf{E}\left[Y\right] = 7/4$ so $\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right] = 21/8$.
Therefore,

$$\mathbf{E}\left[XY\right] \neq \mathbf{E}\left[X\right]\mathbf{E}\left[Y\right].$$

Joint Probability Mass Function

The joint PMF of r.v. X, Y is the function $p_{XY} : \mathbb{R}^2 \to \mathbb{R}$ defined by $p_{XY}(x, y) = \mathbf{Pr}\left[X = x \wedge Y = y\right]$.

Useful equation: With the joint PMF of r.v. X, Y you can compute the expectation of any function $f(X, Y)$:

$$\mathbf{E}\left[f(X, Y)\right] = \sum_{x,y} f(x, y) \cdot p_{XY}(x, y).$$

Compute $\mathbf{E}\left[\frac{X}{Y}\right]$ for the previous r.v. X, Y

$$\mathbf{E}\left[\frac{X}{Y}\right] = p_{XY}(1,1)\frac{1}{1} + p_{XY}(1,2)\frac{2}{1} + p_{XY}(1,3)\frac{3}{1} + p_{XY}(2,1)\frac{1}{2} + p_{XY}(2,2)\frac{2}{2} + p_{XY}(3,2)\frac{2}{3} = \frac{5}{4}$$

Independent r.v.

Two random variables X and Y are said to be independent if

$$\forall x, y \in \mathbb{R}, \mathbf{Pr}\left[(X = x) \cap (Y = y)\right] = \mathbf{Pr}\left[X = x\right] \cdot \mathbf{Pr}\left[Y = y\right].$$

Two not independent r.v. are said to be correlated.

Rolling two dices, let X_1 be a r.v. counting the pips in dice 1, and let X_2 be a r.v. counting the pips in dice 2. Then X_1 and X_2 are independents.
Rolling two dices, let X_1 be a r.v. counting the pips in dice 1, and let X_3 count the sum of pips in the two rollings, then X_1 and X_3 are correlated.

Independent r.v.: Main result

Theorem If X and Y are independent r.v. then
$\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$.
Proof

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y} p_{XY}(x,y) \cdot xy \\
&= \sum_{x,y} p_X(x) p_Y(y) \cdot xy \text{ (by independence)} \\
&= \sum_{x,y} x p_X(x) y p_Y(y) = \left(\sum_x x p_X(x)\right)\left(\sum_y y p_Y(y)\right) \\
&= \mathbf{E}[X]\,\mathbf{E}[Y] \qquad \square
\end{aligned}$$

Recall that if X and Y are independent, then for any real value f and g, $f(X)$ and $g(Y)$ also are independent
$\Rightarrow \mathbf{E}[f(X) \cdot g(Y)] = \mathbf{E}[f(X)] \cdot \mathbf{E}[g(Y)]$

The Poisson approximation to the Binomial

For $X \in B(n, p)$, for large n, computing the PMF $\mathbf{Pr}[X = x]$ could be quite nasty.
It turns out that for large n and small p, $B(n, p)$ can be easily approximated by the PMF of a simpler Poisson random variable.

A discrete r.v. X is Poisson with parameter λ ($X \in P(\lambda)$), if it has PMF $\mathbf{Pr}[X = i] = \frac{\lambda^i e^{-\lambda}}{i!}$, for $i \in \{0, 1, 2, 3, \ldots\}$

If $X \in P(\lambda)$ then $\mathbf{E}[X] = \lambda$.

This is the reason that sometimes λ is denoted μ.

Proof:

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} i \frac{\lambda^i e^{-\lambda}}{i!} = e^{-\lambda}\lambda \underbrace{\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}}_{\text{Taylor for } e^{\lambda}} = e^{-\lambda}\lambda e^{\lambda} = \lambda \qquad \square$$

The Poisson approximation to the Binomial

Theorem: If $X \in B(n,p)$, with $\mu = pn$, then as $n \to \infty$, for each fixed $i \in \{0,1,2,3,\ldots\}$,

$$\mathbf{Pr}\left[X = i\right] \sim \frac{\mu^i e^{-\mu}}{i!}.$$

Proof: As $\mu = np$,

$$\begin{aligned}
\mathbf{Pr}\left[X = i\right] &= \binom{n}{i}(\frac{\mu}{n})^i(1-\frac{\mu}{n})^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{i!}\frac{\mu^i}{n^i}(1-\frac{\mu}{n})^n(1-\frac{\mu}{n})^{-i} \\
&= \frac{\mu^i}{i!}(1-\frac{\mu}{n})^n\frac{n(n-1)\cdots(n-i+1)}{n^i}(1-\frac{\mu}{n})^{-i} \\
&\sim \frac{\mu^i}{i!}e^{-\mu} \text{ as } n \to \infty. \qquad \square
\end{aligned}$$

Example

The population of Catalonia is around 7 million people. Assume Suppose that the probability that a person is killed by lightning in a year is, independently, $p = \frac{1}{5\times 10^8}$.

a.- Compute the exact probability that 3 or more people will be killed by lightning next year in Catalonia.

Let X be a r.v. counting the number of people that will be killed in Cat. next year by a lightning.

We want to compute

$\mathbf{Pr}[X \geq 3] = 1 - \mathbf{Pr}[X \geq 0] - \mathbf{Pr}[X = 1] - \mathbf{Pr}[X = 2]$, where $X \in B(7 \times 10^6, \frac{1}{5\times 10^8})$.

Then,

$\mathbf{Pr}[X \geq 3] = 1 - (1-p)^n - np(1-p)^{n-1} - \binom{n}{2} p^2 (1-p)^{n-2} = 1.65422 \times 10^{-7}$

Example

b.- Approximate $\mathbf{Pr}[X \geq 3]$ $\lambda = np = 7/500$ so
$\mathbf{Pr}[X \geq 3] \sim 1 - e^{\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} = 1.52558 \times 10^{-7}$

c.- Approximate the probability that 2 or more people will be killed by lightning the first 6 months of 2019
Notice we are considering λ as a *rate*. Then $\lambda = 7/2 \times 500$
$\mathbf{Pr}[X \geq 2 \text{ during 6 months}] \sim 1 - e^{\lambda} - \lambda e^{-\lambda} = 5.79086 \times 10^{-7}$

d.- Approximate the probability that in 3 of the next 10 years exactly 3 people will be killed
We have $\lambda = 7/500$, then the probability that every year 3 people are killed $= \frac{e^{-\lambda}\lambda^3}{3!}$. Let Y be a r.v. counting the number of years with exactly 3 kills.
Assuming independence between years, $Y \in B(19, \frac{e^{-\lambda}\lambda^3}{3!})$, therefore the answer is $\binom{10}{3}(\frac{e^{-\lambda}\lambda^3}{3!})^3(1 - \frac{e^{-\lambda}\lambda^3}{3!})^7$