Topic 3: The Expectation of a Random Variable

Rohini Somanathan

Course 003, 2018

Expectation of a discrete random variable

Definition (Expectation of a discrete r.v.): *The expected value (also called the expectation or mean) of a discrete random variable whose distinct possible values are* $x_1, x_2, \ldots$ *is defined by*

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

This is a probability-weighted average of possible values of X. Think of the expectation as a point of balance: if weights are placed on a weightless rod, where should a fulcrum be placed so that the rod balances?

Note:

- **$E(X)$ need not be in the support of X:**

  If X = number of dots in the roll of a die. $E(X) = \sum_{x=1}^{6} \frac{x}{6} I_{\{1,2\ldots6\}}(x) = 3.5$

- **If the support of X is finite, this expectation always exists. If infinite, the probability of high values of X must be small.**

Expectations: some useful results

- **Result 1:** If X and Y are r.v.'s with the same distribution, then

$$E(X) = E(Y)$$

(if either side exists)

- **Result 2:** For any r.v. X and a constant c

$$E(cX) = cE(X)$$

- **Result 3 (linearity of expectation):** For *any* r.v.s X, Y,

$$E(X + Y) = E(X) + E(Y)$$

Note: We do not require independence of X and Y.

Example: 2 coins, X = at least one head, Y = at least one tail- verify linearity of expectations using the distributions of X, Y and $X + Y$.

Let the events $\{s\}$ denote the smallest events in the sample space over which X and Y are defined. Then

$$E(X) + E(Y) = \sum_s X(s)P(\{s\}) + \sum_s Y(s)P(\{s\}) = \sum_s (X+Y)(s)P(\{s\}) = E(X+Y)$$

The expectations of Bernoulli and Binomial r.v.s

Bernoulli: Using the definition of the expectation, this is $1p + 0q = p$

Binomial: A binomial r.v. can be expressed as the sum of n Bernoulli r.v.s:

$$X = I_1 + I_2 + \dots I_n.$$

By linearity,

$$E(X) = E(I_1) + E(I_2) + \dots E(I_n) = np.$$

The Geometric distribution

Definition (Geometric distribution): *Consider a sequence of independent Bernoulli trials, all with success probability $p \in (0,1)$. Let X be the number of failures before the first successful trial. Then $X \sim Geom(p)$*

The PMF is given by

$$P(X = k) = q^k p$$

for $k = 0, 1, 2, \ldots$ where $q = 1 - p$. This is a valid PMF since the sum $p \sum\limits_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1$

The expectation of a geometric r.v. is defined as

$$E(X) = \sum_{k=0}^{\infty} k q^k p$$

We know that $\sum\limits_{k=0}^{\infty} q^k = \frac{1}{1-q}$. Differentiating both sides w.r.t. q, we get $\sum\limits_{k=0}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$. Now multiplying both sides by pq, we get

$$E(X) = pq \frac{1}{(1-q)^2} = \frac{q}{p}$$

The Negative Binomial distribution

Definition (Negative Binomial distribution): *Consider a sequence of independent Bernoulli trials, all with success probability* $\mathbf{p} \in (0,1)$*. Let* $\mathbf{X}$ *be the number of failures before the* $\mathbf{rth}$ *success. Then* $\mathbf{X} \sim NBin(\mathbf{r}, \mathbf{p})$

The PMF is given by

$$P(X = k) = \binom{k + r - 1}{r - 1} p^r q^k$$

for $k = 0, 1, 2, \ldots$ **where** $q = 1 - p$**.**

We can write the Negative Binomial as a sum of r **Geom**(p) **r.v.s:** $X = X_1 + X_2 \cdots + X_r$

The expectation of a Negative Binomial r.v. is therefore

$$E(X) = r\frac{q}{p}$$

Expectation of a function of X

Result: *If X is a discrete random variable and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then*

$$E(g(X)) = \sum_x g(x) P(X = x)$$

where the sum is taken over all possible values of X.

Intuition: **Whenever $X = x$, $g(X) = g(x)$, so we can assign $p(x)$ to $g(x)$. An easy but illustrative example is $g(X) = X^3$**

This is a very useful result, because it tells us we don't need the PMF of $g(X)$ to find its expected value, and we are often interested in functions of random variables.

Examples: **expected revenues from the distribution of yields, earnings from a chance game...**

Variance of a random variable

Definition (variance and standard deviation): *The variance of and r.v. X is*

$$Var(X) = E[(X - EX)^2]$$

The square root of the variance is called the standard deviation of X.

$$SD(X) = \sqrt{Var(X)}$$

Notice that by linearity we have $E(X - EX) = 0$

We can expand the above expression to rewrite the variance in a form that is often more convenient:

$$Var(X) = E(X^2) - (EX)^2$$

Variance properties

1. $Var(X) \geq 0$ with equality if and only if $P(X = a) = 1$ for some constant a.

2. $Var(aX + b) = a^2 Var(X)$ for any constants a and b. It follows that $Var(X) = Var(-X)$

   Proof: $Var(aX + b) = E[(aX + b - (aEX + b))^2] = E[(a(X - EX))^2] = a^2 E[(X - EX)^2] = a^2 Var(X)$

3. $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$ for independent random variables $X_1, \dots, X_n$.

   Proof: Denote EX by μ. For $n = 2$, $E(X_1 + X_2) = \mu_1 + \mu_2$ and therefore

$$Var(X_1 + X_2) = E[(X_1 + X_2 - \mu_1 - \mu_2)^2] = E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)]$$

   Taking expectations, we get
   $E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] = Var(X_1) + Var(X_2) + 2E[(X_1 - \mu_1)(X_2 - \mu_2)]$
   But since X_1 and X_2 are independent,

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = E(X_1 - \mu_1)E(X_2 - \mu_2) = (\mu_1 - \mu_1)(\mu_2 - \mu_2) = 0$$

   It therefore follows that

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$

   This holds for the sum of any number of independent random variables.

The Poisson distribution

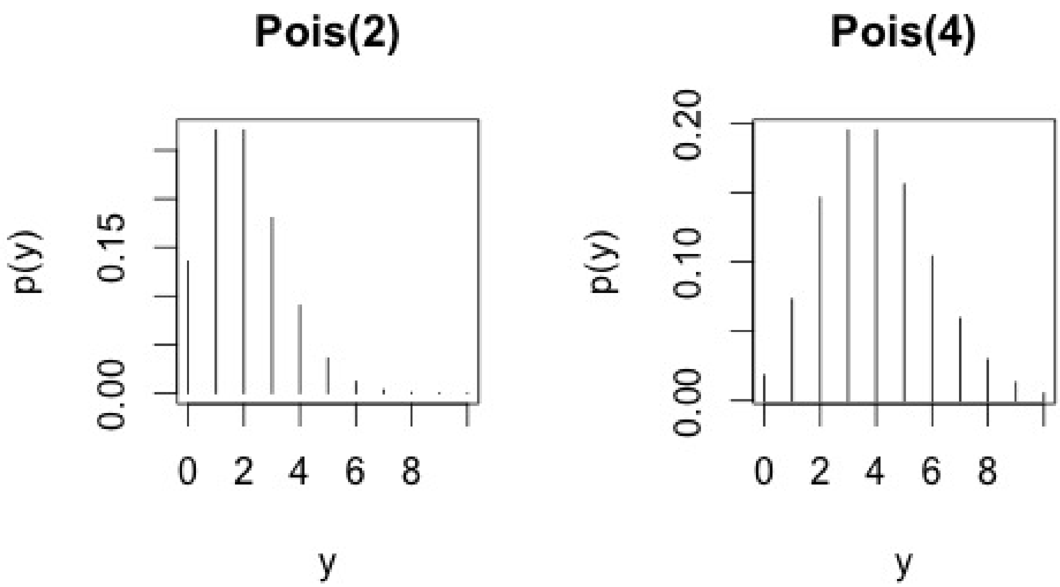
Definition (Poisson distribution): *An r.v.* $\mathbf{X}$ *has the Poisson distribution with parameter* $\boldsymbol{\lambda} > \mathbf{0}$ *if the PMF of X is*

$$\mathbf{P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}}$$

for $\mathbf{k = 0, 1, 2, \ldots}$. *We write this as* $\mathbf{X \sim Pois(\lambda)}$.

This is a valid PMF because the Taylor series $\mathbf{1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots}$ **converges to** $\mathbf{e^{\lambda}}$ **so** $\mathbf{\sum\limits_k f(k) = \sum\limits_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1}$.

As $\boldsymbol{\lambda}$ **gets larger, the PMF becomes more bell-shaped. The mean and variance are both** $\boldsymbol{\lambda}$.



Poisson additivity

Result: *If* $X \sim Pois(\lambda_1)$ *and* $Y \sim Pois(\lambda_2)$ *are independent, then* $X + Y \sim Pois(\lambda_1 + \lambda_2)$

Proof: **To get the PMF of $X + Y$, we use the law of total probability:**

$$\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X+Y=k|X=j)P(X=j) \\
&= \sum_{j=0}^{k} P(Y=k-j)P(X=j) \\
&= \sum_{j=0}^{k} \frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!}\frac{e^{-\lambda_1}\lambda_1^{j}}{j!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\lambda_1^j\lambda_2^{k-j} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!} \text{ (using the binomial theorem)}
\end{aligned}$$

A Poisson example

A binomial distribution with large n and small p can be approximated by a Poisson, which is computationally much easier.

We have a 300 page novel with $1,500$ letters on each page.

Typing errors are as likely to occur for one letter as for another, and the probability of such an error is given by $p = 10^{-5}$.

The total number of letters $n = (300) * (1500) = 450,000$

Using $\lambda = np$, the poisson distribution function gives us the probability of the number of errors being less than or equal to 10 as:

$$P(x \leq 10) \approx \sum_{x=0}^{10} \frac{e^{-4.5}(4.5)^x}{x!} = .9933$$

Rules of Thumb: close to binomial probabilities when $n \geq 20$ and $p \leq .05$, excellent when $n \geq 100$ and $np \leq 10$.