50.034 - Introduction to Probability and Statistics

Week 3 – Lecture 6

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 6 Assigned Readings: 5.1, 5.2, 5.4, 5.5, 5.7

Outline of Lecture

- Properties of variance
- Bernoulli and Binomial distributions
- Geometric distribution
- Poisson distribution
- Exponential distribution

Variance of a linear function

Recall: The variance of *any* R.V. X is $\mathrm{var}(X) = \mathbf{E}[(X - \mu_X)^2]$, provided that $\mathbf{E}[(X - \mu_X)^2]$ exists.

Theorem: Let X be *any* R.V., and let a and b be finite constants. Then,

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X).$$

Proof:

$$\begin{aligned}
\mathrm{var}(aX + b) &= \mathbf{E}\left[(aX + b - \mathbf{E}[aX + b])^2\right] \\
&= \mathbf{E}\left[(aX + b - a\mathbf{E}[X] - b)^2\right] \\
&= \mathbf{E}\left[(aX - a\mathbf{E}[X])^2\right] \\
&= \mathbf{E}\left[a^2(X - \mathbf{E}[X])^2\right] \\
&= a^2\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\
&= a^2 \mathrm{var}(X).
\end{aligned}$$

Variance of sum of independent R.V.'s

Recall: Events A, B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

From independent events to independent R.V.'s:

- Two R.V.'s X and Y are independent if for all sets C, C' of real numbers, the events $\{X \in C\}$, $\{Y \in C'\}$ are independent.
- R.V.'s $X_1, \ldots, X_n$ are independent if for all sets $C_1, \ldots, C_n \subseteq \mathbb{R}$, the events $\{X_1 \in C_1\}, \ldots, \{X_n \in C_n\}$ are **mutually independent**.
  [**Recall:** (Lecture 3) Events $A_1, \ldots, A_n$ are mutually independent if the probability of the intersection of any subset of the n events is equal to the product of the individual probabilities.]
- We shall look at independent R.V.'s again in next two lectures.

Theorem: If $X_1, \ldots, X_n$ are **independent** R.V.'s with finite means, then $\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)$.

- Remember: "variance of sum" = "sum of variances", under the assumption that the R.V.'s are **independent**.

Variance of sum of independent R.V.'s

Proof: First, consider the case $n = 2$.
Note: "mean of sum of R.V.'s" = "sum of means of R.V.'s", thus

$$\mathbf{E}[X_1 + X_2] = \mu_{X_1} + \mu_{X_2}.$$

$$\begin{aligned}
\mathrm{var}(X_1 + X_2) &= \mathbf{E}[(X_1 + X_2 - \mu_{X_1} - \mu_{X_2})^2] \\
&= \mathbf{E}[(X_1 - \mu_{X_1})^2 + (X_2 - \mu_{X_2})^2 + 2(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] \\
&= \mathrm{var}(X_1) + \mathrm{var}(X_2) + 2\mathbf{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})].
\end{aligned}$$

Since X_1, X_2 are independent,

$$\begin{aligned}
\mathbf{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] &= \mathbf{E}[(X_1 - \mu_{X_1})]\mathbf{E}[(X_2 - \mu_{X_2})] \\
&= (\mu_{X_1} - \mu_{X_1})(\mu_{X_2} - \mu_{X_2}) \\
&= 0.
\end{aligned}$$

Therefore, $\mathrm{var}(X_1 + X_2) = \mathrm{var}(X_1) + \mathrm{var}(X_2)$.
The general case follows by induction on n.

Recall: Bernoulli distribution (Lecture 4)

A discrete R.V. X is called Bernoulli if it takes two values 0 and 1.

- The pmf $p(x)$ of X is given by $p(1) = p$, $p(0) = 1 - p$ for some $0 \leq p \leq 1$. This pmf is also written often as $p(x; p)$.
- We say that X is the Bernoulli R.V. with parameter p.
- A Bernoulli distribution is the distribution of a Bernoulli R.V.
- (Example 5 of Lecture 5:) $\mathbf{E}[X] = p$, and $\mathrm{var}(X) = p(1 - p)$.

Main Use: To model a **single** trial of a Bernoulli process.

- A Bernoulli process is an experiment with two outcomes.
- e.g. a coin toss has two outcomes: 1 (heads) and 0 (tails).
- A single trial of a Bernoulli process is called a Bernoulli trial.
- The parameter p is usually called the success rate, especially if 1 represents success and 0 represents failure.

Recall: Binomial distribution (Lecture 4)

A discrete R.V. X is called binomial if it takes on finitely many values $0, 1, \ldots, n$, and its pmf $p(x)$ is given by:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x},$$

for some $0 \leq p \leq 1$.

- The pmf $p(x)$ is also written often as $p(x; n, p)$.
- $\binom{n}{x}$ here is a binomial coefficient (from 1st Cohort class).
- We say that X is the binomial R.V. with parameters n and p.
- A binomial distribution is the distribution of a binomial R.V.

Main Use: To model the sum of n trials of a Bernoulli process.

- e.g. the number of heads of n coin tosses.
- The parameter p is usually called the success rate.
- A Bernoulli R.V. is a binomial R.V. with one trial (i.e. $n = 1$).

Mean and variance of binomial R.V.

Proposition: Let X be the binomial R.V. with parameters n and p. Its mean and variance are $\mathbf{E}[X] = np$ and $\mathrm{var}(X) = np(1-p)$.

Proof: By definition $X = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are n *independent* Bernoulli R.V.'s, each with parameter p.
Note: $\mathbf{E}[X_i] = p$ and $\mathrm{var}(X_i) = p(1-p)$ for each $i = 1, \ldots, n$.

Since "mean of sum of R.V.'s" = "sum of means of R.V.'s",

$$\mathbf{E}[X] = \mathbf{E}[X_1 + \cdots + X_n] = \mu_{X_1} + \cdots + \mu_{X_n} = np.$$

Since $X_1, \ldots, X_n$ are independent, "variance of sum" = "sum of variances", i.e.

$$\mathrm{var}(X) = \mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n) = np(1-p).$$

Geometric distribution

A discrete R.V. X is called geometric if it takes on non-negative integer values $0, 1, 2, \ldots,$ and its pmf $p(x)$ is given by:

$$p(x) = p(1-p)^x,$$

for some $0 < p < 1$.

- We say that X is the geometric R.V. with parameter p.
- A geometric distribution is the distribution of a geometric R.V.
- We can verify that

$$\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} p(1-p)^x = p \sum_{i=0}^{\infty} (1-p)^i = 1,$$

  using the **geometric** series $1 + \frac{1}{1-p} + \frac{1}{(1-p)^2} + \cdots = \frac{1}{1-(1-p)}$.

Main Use: To model the number of failed trials in a Bernoulli process immediately before the first success.

- e.g. the number of coin tosses before getting the first heads

Example 1

A weather experiment involves checking if there is rain each day. Suppose the probability of having rain on any given day is 0.3. What is the probability that the first rainy day occurs on the 10th day of the experiment?

Solution:
Let X be the the number of rain-less days, immediately before the first rainy day. Then X can be modeled as a geometric R.V. with parameter $p = 0.3$.

$$\Pr(\text{first rain on 10th day}) = \Pr(X = 9) = 0.3 \cdot (0.7)^9 \approx 0.01211.$$

Properties of geometric distributions

Proposition: Let X be the geometric R.V. with parameter p. Its mean and variance are $\mathbf{E}[X] = \frac{1-p}{p}$ and $\mathrm{var}(X) = \frac{1-p}{p^2}$.

Theorem: Every geometric R.V. X has the memoryless property, i.e. $\Pr(X = k + t | X \geq k) = \Pr(X = t)$ for all integers $k, t \geq 0$.

Intuition: Toto lottery draw with consecutive winning numbers

- Suppose that in the past 1000 draws, the six winning numbers are never consecutive. Does it mean that consecutive winning numbers are "due"? Is the probability of having consecutive winning numbers higher now?

No! The number of failures (non-consecutive winning numbers) follows a **geometric distribution**. (lottery draw = Bernoulli trial.)

- By the **memoryless property**, the probability of having six consecutive winning numbers is exactly the same, whether or not there are 1000 failures or millions of failures.

Recall: Poisson distribution (Lecture 5)

A discrete R.V. X is called Poisson if it takes on non-negative integer values $0, 1, 2, \ldots,$ and its pmf $p(x)$ is given by:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!},$$

for some real number $\lambda > 0$.

- The pmf $p(x)$ is also written often as $p(x; \lambda)$.
- We say that X is the Poisson R.V. with parameter λ.
- A Poisson distribution is the distribution of a Poisson R.V.

Common Use: To model the number of occurrences of an event during a certain time period.

- e.g. number of cars entering SUTD in the morning.
- e.g. number of people visiting Apple Store in the past hour.

Mean and variance of Poisson R.V.

Proposition: Let X be the Poisson R.V. with parameter λ. Its mean and variance are $\mathbf{E}[X] = \lambda$ and $\mathrm{var}(X) = \lambda$.

Proof: We already showed $\mathbf{E}[X] = \lambda$ in Example 2 of Lecture 5. We shall now use a similar "trick" by considering $\mathbf{E}[X(X-1)]$:

$$\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) \cdot p(x;\lambda) = \sum_{x=0}^{\infty} \frac{x(x-1)\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=2}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-2)!} = \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2} e^{-\lambda}}{(x-2)!} = \lambda^2 \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \lambda^2,
\end{aligned}$$

where the last equality follows from the identity:

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=0}^{\infty} p(x;\lambda) = 1.$$

Mean and variance of Poisson R.V.

Proof: (continued)
So far: $\mathbf{E}[X] = \lambda$, and $\mathbf{E}[X(X-1)] = \lambda^2$.

Since "mean of sum" = "sum of means", we get:

$$\mathbf{E}[X(X-1)] = \mathbf{E}[X^2 - X] = \mathbf{E}[X^2] - \mathbf{E}[X] = \mathbf{E}[X^2] - \lambda,$$

hence

$$\mathbf{E}[X^2] = \lambda^2 + \lambda.$$

Now, using the formula for variance,

$$\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

Remark: Since $\mathbf{E}[\text{Poisson R.V. with parameter } \lambda] = \lambda$, we sometimes say "Poisson R.V. with mean λ" instead of "Poisson R.V. with parameter λ".

Question

There are 100 products. The probability that a product has flaws is $p = 0.01$. What is the probability that only 1 product has flaws?

Solution:
Let X be the discrete R.V. representing the number of products with flaws. X can be modeled as a binomial R.V. with parameters $n = 100$ and $p = 0.01$.

- **Note:** p does not always have to be an actual "success" rate. It could be the rate of interest, which in this case is the rate that a product has flaws.

We can use a calculator to get:

$$\Pr(X = 1) = \binom{100}{1} p^1 (1-p)^{99} = 0.3697.$$

Question

What is the probability that at most 3 products have flaw?

Solution: Same as before, let X be a binomial R.V. with parameters $n = 100$ and $p = 0.01$.

$$\begin{aligned}\Pr(X \leq 3) = \Pr(X = 0) + \Pr(X = 1) + \\ \Pr(X = 2) + \Pr(X = 3) \\ = \binom{100}{0}(1-p)^{100} + \binom{100}{1}p^1(1-p)^{99} + \\ \binom{100}{2}p^2(1-p)^{98} + \binom{100}{3}p^3(1-p)^{97}\end{aligned}$$

Can you get the answer by using a calculator?

Yes, but the numerical calculation becomes more complicated.

- It is tedious to calculate values from a binomial distribution.

Poisson distribution as a limit

Theorem: Let X be a binomial R.V. with parameters n and p. Suppose we let $n \to \infty$ and $p \to 0$ in such a way that $np \to \lambda$ for some value $\lambda > 0$. Then the pmf of X approaches the pmf of the Poisson R.V. with parameter λ, i.e. $p(x; n, p) \to p(x; \lambda)$.

Corollary: For any binomial distribution in which n is large and p is small, $p(x; n, p) \approx p(x; \lambda)$, where $\lambda = np$.

Rule of Thumb: Poisson distribution is a good approximation to binomial distribution if $n > 50$ and $np < 5$.

Back to the previous question

In the question, we have $n = 100$ and $p = 0.01$.

Since $n > 50$ and $np < 5$, we can use a Poisson distribution to approximate the binomial distribution.

The parameter λ of the Poisson distribution is $np = 1$; therefore the pmf can be approximated by $p(x; 1) = \dfrac{e^{-1}}{x!}$, for $x = 0, 1, \ldots$

$$\begin{aligned}\Pr(X \leq 3) &= \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) \\ &= e^{-1}(1 + 1 + \tfrac{1}{2} + \tfrac{1}{6}) \approx 0.9810\end{aligned}$$

If we carried out the tedious binomial pmf computation, we would actually get $\Pr(X \leq 3) \approx 0.9816$.

- Not bad! Poisson distribution gives a good approximation.

Example 2

Suppose X is a binomial R.V. with parameter n and p, such that $n > 100$, $np < 5$, and in particular, $np = \ln 2$.

Which of the following options is a good approximation to the probability that $X = 0$?

A) 0.3

B) 0.5

C) 0

Answer: B.
$\Pr(X = 0) = p(0; n, p) \approx p(0; \ln 2) = e^{-\ln 2} = 0.5$.
(Recall: $p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$.)

Recall: Exponential distribution (Lecture 4)

A continuous R.V. X is called exponential if its pdf is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

for some real number $\lambda > 0$.

- The pdf $f(x)$ is also written often as $f(x; \lambda)$.
- We say that X is the exponential R.V. with parameter λ.
- Its distribution is called exponential distribution.
- Its cdf is $F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$.

Common Use: To model the elapsed time between two successive events.

- e.g. time between two successive cars entering SUTD.
- e.g. time between two people visiting Apple Store.

Mean and variance of exponential R.V.

Proposition: Let X be the exponential R.V. with parameter λ. Its mean and variance are $\mathbf{E}[X] = \frac{1}{\lambda}$ and $\mathrm{var}(X) = \frac{1}{\lambda^2}$.

Proof: Using integration by parts,

$$\begin{aligned}
\mathbf{E}[X] &= \int_{x=0}^{\infty} x\lambda e^{-\lambda x}\,dx = \Big[-xe^{-\lambda x}\Big]_{x=0}^{x=\infty} - \int_0^\infty (-e^{-\lambda x})\,dx \\
&= [0-0] - \Big[\tfrac{1}{\lambda}e^{-\lambda x}\Big]_{x=0}^{x=\infty} = \tfrac{1}{\lambda}. \\
\mathbf{E}[X^2] &= \int_{x=0}^{\infty} x^2\lambda e^{-\lambda x}\,dx = \Big[-x^2e^{-\lambda x}\Big]_{x=0}^{x=\infty} - \int_0^\infty (-2xe^{-\lambda x})\,dx \\
&= [0-0] + \tfrac{2}{\lambda}\cdot\int_0^\infty x\lambda e^{-\lambda x}\,dx = \tfrac{2}{\lambda^2} \text{ (from previous eqn).}
\end{aligned}$$

Thus, the variance is

$$\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \tfrac{2}{\lambda^2} - \left(\tfrac{1}{\lambda}\right)^2 = \tfrac{1}{\lambda^2}.$$

Poisson and Exponential distributions

The two distributions are profoundly related.

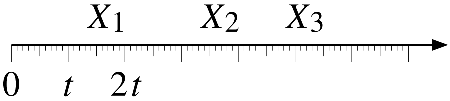
Suppose the number of events occurring in a time interval t has a Poisson distribution with parameter $\lambda_{Poisson} = \alpha t$, where α is the number of occurrences per hour (i.e. α is the occurence rate).

Goal: We want to study the probability that the waiting time for the first event is not more than t.

- Although the waiting time is a continuous R.V., we want to calculate the probability based on Poisson distribution.

Poisson and Exponential distributions

Here is how we do: Let $T_1, T_2, \ldots$ be R.V.'s such that each T_k represents the time at which the k-th event occurs.



The probability of the waiting time for the first event being not more than t is $\Pr(T_1 \leq t)$.

$$\Pr(T_1 \leq t) = 1 - \Pr(T_1 > t)$$

$\Pr(T_1 > t) =$ probability that no event occurs in the interval $[0, t]$.

Poisson and Exponential distributions

Now we can apply the pmf $p(x; \lambda_{Poisson}) = p(x; \alpha t)$ corresponding to the Poisson distribution, since the time interval is clearly defined.

$$\begin{aligned}
\Pr(T_1 \leq t) &= 1 - \Pr(T_1 > t) \\
&= 1 - \Pr(\text{no event in } [0, t]) \\
&= 1 - \frac{\lambda_{Poisson}^0 e^{-\lambda_{Poisson}}}{0!} \\
&= 1 - e^{-\alpha t}
\end{aligned}$$

This gives exactly the cdf of the exponential distribution!

Poisson and Exponential distributions

Taking the derivative with respect to t, we obtain

$$f(t) = \frac{d}{dt}(1 - e^{-\alpha t}) = \alpha e^{-\alpha t}$$

By comparing this expression with the pdf $f(x; \lambda_{exponential})$ of an exponential R.V., we see that $x = t$ and $\lambda_{exponential} = \alpha$.

Conclusion: The occurrence rate α for the Poisson distribution is the parameter of the exponential distribution.

- Since $\mathbf{E}$[Poisson R.V. with parameter $\lambda] = \lambda$, the value of $\lambda_{Poisson}$ is the average number of occurrences.
- Since $\lambda_{Poisson} = \alpha t$, we can determine the occurrence rate α once we know $\lambda_{Poisson}$ (average number of occurrences) and t (length of time interval).

Example 3

Customers arrive in a shop following a Poisson distribution at an average rate of 20 customers per hour. What is the probability that the shopkeeper will have to wait for more than 5 minutes for the arrival of the first customer?

Solution:
Let T be the R.V. representing the waiting time in minutes until the first customer arrives. T can be modeled as an exponential R.V. with parameter $\lambda = \frac{20}{60} = \frac{1}{3}$ (i.e. the occurrence rate is $\frac{1}{3}$).
The pdf of T is

$$f(t) = \begin{cases} \frac{1}{3}e^{-t/3}, & \text{if } t \geq 0; \\ 0, & \text{if } t < 0; \end{cases}$$

therefore $\Pr(T > 5) = \int_5^\infty \frac{1}{3}e^{-t/3}\,dt = e^{-5/3} \approx 0.1889.$

Example 4

The R.V. X, representing the life of a type of bulb, follows an exponential distribution with mean life of 500 hours. Its pdf is

$$f(x) = \begin{cases} \frac{1}{500} e^{-x/500}, & \text{if } x \geq 0; \\ 0, & \text{otherwise;} \end{cases}$$

Note: The parameter of the exponential R.V. X is $\frac{1}{500}$, not 500, since $\mathbf{E}[\text{exponential R.V. with parameter } \lambda] = \frac{1}{\lambda}$.

What is the probability that a bulb can last for 600 hours?

Solution:

$$\Pr(X \geq 600) = \int_{600}^{\infty} \frac{1}{500} e^{-x/500} \, dx = e^{-1.2} \approx 0.3012.$$

Example 4

Suppose we are given that a bulb has worked for 300 hours. What is the probability that the bulb can last for another 600 hours?

Solution: The conditional probability is

$$\begin{aligned}\Pr(X \geq 900 | X \geq 300) &= \frac{\Pr(X \geq 900)}{\Pr(X \geq 300)} = \frac{e^{-900/500}}{e^{-300/500}} \\ &= e^{-1.2}.\end{aligned}$$

In other words, the probability that the bulb can last for another 600 hours remains the same, whether or not we are given the conditioning event that the bulb has worked for 300 hours.

- More generally, every exponential R.V. X satisfies $\Pr(X \geq t + h | X \geq t) = \Pr(X \geq h)$ for all $t > 0$, $h > 0$.
- In other words, every exponential distribution has the memoryless property.

μ and σ^2 of several probability distributions

Bernoulli distribution (with parameter p) *[discrete R.V.]*

$$\mu = p \text{ and } \sigma^2 = p(1-p)$$

Binomial distribution (with parameters n and p) *[discrete R.V.]*

$$\mu = np \text{ and } \sigma^2 = np(1-p)$$

Geometric distribution (with parameter p) *[discrete R.V.]*

$$\mu = \frac{1-p}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}$$

Poisson distribution (with parameter λ) *[discrete R.V.]*

$$\mu = \lambda \text{ and } \sigma^2 = \lambda$$

Exponential distribution (with parameter λ) *[continuous R.V.]*

$$\mu = \frac{1}{\lambda} \text{ and } \sigma^2 = \frac{1}{\lambda^2}$$

Summary

- Properties of variance
- Bernoulli and Binomial distributions
- Geometric distribution
- Poisson distribution
- Exponential distribution

Reminder:
There is **mini-quiz 1** (15mins) this week during Cohort Class!

- Tested on materials from Lecture 1 up to and including Slide 7 ("Mean and variance of binomial R.V.") of today's lecture.