Week 05 notes

Discrete families: summary

01 Theory

⚠ Memorize this info!

Bernoulli: $X \sim \text{Ber}(p)$

$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$

$\text{Var}[X] = E[X^2] - E[X]^2$

$E[X^2] = 1 \cdot p + 0 \cdot (1-p) = p$

$\text{Var} = p - p^2 = p(1-p) = pq$

- Indicates a win.
- $P_X(1) = p, \;\; P_X(0) = q$
- $E[X] = p$
- $\text{Var}[X] = pq$

Binomial: $X \sim \text{Bin}(n, p)$

- Counts number of wins.
- $P_X(k) = \binom{n}{k} p^k q^{n-k}$
- $E[X] = np$
- $\text{Var}[X] = npq$
- These are n times the Bernoulli numbers.

Geometric: $X \sim \text{Geom}(p)$

- Counts discrete wait time until first win.
- $P_X(k) = q^{k-1} p$
- $E[X] = \frac{1}{p}$
- $\text{Var}[X] = \frac{q}{p^2}$

Pascal: $X \sim \text{Pasc}(\ell, p)$

- Counts discrete wait time until ℓ^{th} win.
- $P_X(k) = \binom{k-1}{\ell-1} q^{k-\ell} p^\ell$
- $E[X] = \frac{\cancel{k}\,\ell}{p}$
- $\text{Var}[X] = \frac{\cancel{k}\,\ell\, q}{p^2}$
- These are k times the Geometric numbers.

Poisson: $X \sim \text{Pois}(\lambda)$

- Counts "arrivals" during time interval.
- $P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$
- $E[X] = \lambda$
- $\text{Var}[X] = \lambda$

Function on a random variable

02 Theory

By composing any function $g : \mathbb{R} \to \mathbb{R}$ with a random variable $X : S \to \mathbb{R}$ we obtain a new random variable $g \circ X$. The new one is called a **derived** random variable.

- Write $g(X)$ for this derived random variable $g \circ X$.

Expectation of derived variables

Discrete case:

$$E\big[g(X)\big] = \sum_k g(k) \cdot P_X(k)$$

(Here the sum is over all possible values k of X.)

Continuous case:

$$E\big[g(X)\big] = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x)\, dx$$

- Notice: when applied to outcome $s \in S$:
    - k is the output of X
    - $g(k)$ is the output of $g \circ X$

The proofs of these formulas are *not trivial*, since one must relate the PDF or PMF of X to that of $g(X)$.

Proof - Discrete case - Expectation of derived variable

$$E[g(X)] = \sum_y y \cdot P_{g(X)}(y)$$

$P_{g(X)}(y)$
$= P[g(X) = y]$

$$= \sum_y y \cdot \sum_{k \in g^{-1}(y)} P_X(k)$$

$= P[X \in g^{-1}(y)]$

$$= \sum_y \sum_{k \in g^{-1}(y)} g(k) \cdot P_X(k)$$

$$= \sum_k g(k) \cdot P_X(k)$$

Linearity of expectation

For constants a and b:

$$E[aX + b] = aE[X] + b$$

For any X and Y on the same probability model:

$$E[X + Y] = E[X] + E[Y]$$

Exercise - Linearity of expectation

Using the definition of expectation, verify both linearity formulas for the discrete case.

$E[X \cdot X] \neq E[X] \cdot E[X]$

⚠ Be careful!

Usually $E[X^2] \neq E[X]^2$ and $E\big[g(X)\big] \neq g(E[X])$.

Pulling out a and b across E only works because they are constants.

Variance squares the scale factor

For constants a and b:

$\sigma[aX+b] = |a|\sigma[X]$ $\sqrt{a^2} = |a|$

$$\mathrm{Var}[aX+b] = a^2\,\mathrm{Var}[X]$$

Thus variance *ignores the offset* and *squares the scale factor*. It is not linear!

Proof - Variance squares the scale factor

$$\begin{aligned}
\mathrm{Var}[aX+b] &= E[(aX+b-E[aX+b])^2] \\
&= E[(aX+b-a\mu-b)^2] \\
&= E[(aX-a\mu)^2] \\
&= E[a^2(X-\mu)^2] \\
&= a^2\,E[(X-\mu)^2] \\
&= a^2\,\mathrm{Var}[X]
\end{aligned}$$

Extra - Moments

The n^{th} **moment** of X is defined as the expectation of X^n:

Discrete case:

$$E[X^n] = \sum_k k^n \cdot p(k)$$

Continuous case:

$$E[X^n] = \int_{-\infty}^{+\infty} x^n \cdot f(x)\,dx$$

A **central moment of X** is a moment of the variable $X - E[X]$:

$$E\big[(X-E[X])^n\big]$$

The data of all the moments collectively determines the probability distribution. This fact can be very useful! In this way moments give an analogue of a series representation, and are sometimes more useful than the PDF or CDF for encoding the distribution.

03 Illustration

Example - Function given by chart

One to one

Suppose that $g : \mathbb{R} \to \mathbb{R}$ in such a way that $g : 1 \mapsto 4$ and $g : 2 \mapsto 1$ and $g : 3 \mapsto 87$.

$X:$	**1**	**2**	**3**
$P_X(k):$	1/7	2/7	4/7
$Y:$	4	1	87

(or e.g. $g(x) = x^3$)

Then:

$$E[X] = 1 \cdot \frac{1}{7} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{4}{7} \quad \gg\gg \quad \frac{17}{7}$$

And:

$$E[Y] = 4 \cdot \frac{1}{7} + 1 \cdot \frac{2}{7} + 87 \cdot \frac{4}{7} \quad \gg\gg \quad \frac{267}{7}$$

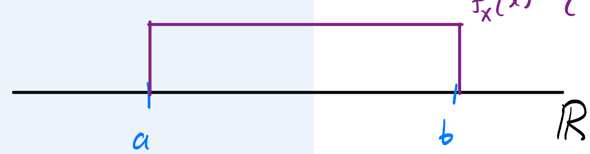
Therefore: $5 \cdot E[X] + 2E[Y] + 3$

$$E[5X + 2Y + 3] \quad \gg\gg \quad 5 \cdot \frac{17}{7} + 2 \cdot \frac{267}{7} + 3 \quad \gg\gg \quad \frac{640}{7}$$

Variance of uniform random variable

The uniform random variable X on $[a, b]$ has distribution given by $P\big[X \in [c, d]\big] = \frac{d-c}{b-a}$ for $[c, d] \subset [a, b]$.

$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{else} \end{cases}$



- (a) Find $\text{Var}[X]$. $= E[X^2] - E[X]^2$
- (b) Find $\text{Var}[3X]$ using "squaring the scale factor."
- (c) Find $\text{Var}[3X]$ directly.

Solution

(a)

1. Compute density.
   - The density for X is:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$\text{Var} = E[X^2] - E[X]^2$

2. Compute $E[X]$ and $E[X^2]$ directly using integral formulas.
   - Compute $E[X]$:

$$E[X] = \int_a^b \frac{x}{b-a}\, dx = \frac{b+a}{2}$$

$$\frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

- Now compute $E[X^2]$:

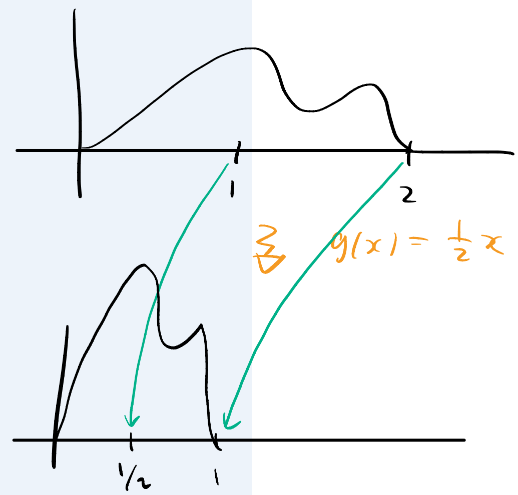
$$E[X^2] = \int_a^b \frac{x^2}{b-a}\,dx \quad \gg\gg \quad \frac{1}{3}(b^2+ba+a^2)$$

3. ≡ Find variance using short formula.
   - Plug in:

$$\mathrm{Var}[X] = E[X^2] - E[X]^2$$

$$\gg\gg \quad \frac{1}{3}(b^2+ab+a^2) - \left(\frac{b+a}{2}\right)^2$$

$$\gg\gg \quad \frac{(b-a)^2}{12}$$



(b)

- "Squaring the scale factor" formula:

$$\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$$

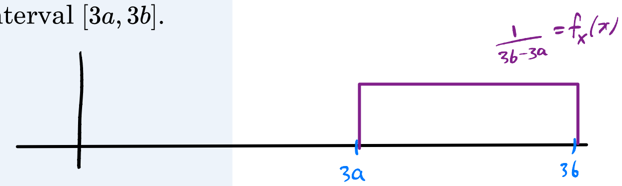
- Plugging in:

$$\mathrm{Var}[3X] \quad \gg\gg \quad 9\mathrm{Var}[X] \quad \gg\gg \quad \frac{9}{12}(b-a)^2$$

(c)

1. ≡ Density.
   - The variable $3X$ will have $1/3$ the density spread over the interval $[3a, 3b]$.
   - Density is then:

$$f_X(x) = \begin{cases} \frac{1}{3b-3a} & \text{on } [3a, 3b] \\ 0 & \text{otherwise} \end{cases}$$



2. ≡ Plug into prior variance formula.
   - Use $a \rightsquigarrow 3a$ and $b \rightsquigarrow 3b$.
   - Get variance:

$$\mathrm{Var}[3X] = \frac{(3b-3a)^2}{12}$$

   - Simplify:

$$\gg\gg \quad \frac{(3(b-a))^2}{12} \quad \gg\gg \quad \frac{9}{12}(b-a)^2$$

Exercise - Probabilities via CDF

Suppose the CDF of X is given by $F_X(x) = \frac{1}{1+e^{-x}}$. Compute:

- (a) $P[X \leq 1]$

- (b) $P[X < 1]$
- (c) $P[-0.5 \leq X \leq 0.2]$
- (d) $P[-2 \leq X]$

Solution

04 Theory

Suppose we are given the PDF $f_X(x)$ of X, a continuous RV.

What is the PDF $f_{g(X)}$, the PDF of the derived variable given by composing X with $g : \mathbb{R} \to \mathbb{R}$?

⚠ PDF of derived

The PDF of $g(X)$ is *not* (usually) equal to $g \circ f_X(x)$.

Relating PDF and CDF

When the CDF of X is differentiable, we have:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad \Longrightarrow \qquad f_X(x) = \frac{d}{dx} F_X(x)$$

$$F_{g(X)}(x) = \int_{-\infty}^{x} f_{g(X)}(t)\,dt \qquad \Longrightarrow \qquad f_{g(X)}(x) = \frac{d}{dx} F_{g(X)}(x)$$

Therefore, if we know $f_X(x)$, we can find $f_{g(X)}(x)$ using a 3-step process:

1. Find $F_X(x)$, the CDF of X, by integration.
   - Compute $\int_{-\infty}^{x} f_X(t)\,dt$.
   - Now remember that $F_X(x) = P[X \leq x]$.
2. Find $F_{g(X)}(x)$, the CDF of $g(X)$, by direct comparison to $F_X(x)$.
   - Remember definition $F_{g(X)}(x) = P[g(X) \leq x]$.
   - Compare $P[g(X) \leq x]$ to $P[X \leq g^{-1}(x)]$, for example.
   - These are equal if g is monotone increasing.
3. Find $f_{g(X)}(x)$, the PDF of $g(X)$, by differentiation.
   - Use $f_X(x) = \frac{d}{dx} F_{g(X)}(x)$.

05 Illustration

Example - PDF of derived from CDF

Suppose that $F_X(x) = \frac{1}{1+e^{-x}}$.

- (a) Find the PDF of X.
- (b) Find the PDF of e^X.

Solution

(a)

- Formula:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad \Longrightarrow \qquad f_X(x) = \frac{d}{dx} F_X(x)$$

- Plug in:

$$f_X(x) = \frac{d}{dx}\left(1+e^{-x}\right)^{-1} \quad \gg\gg \quad -\left(1+e^{-x}\right)^{-2} \cdot (-e^{-x})$$

$$\gg\gg \quad \frac{e^{-x}}{(1+e^{-x})^2}$$

(b)

- By definition:

$$F_{e^X}(x) = P[e^X \leq x]$$

- Since e^X is increasing, we know:

$$e^X \leq a \qquad \Longleftrightarrow \qquad X \leq \ln a$$

- Therefore:

$$F_{e^X}(x) = F_X(\ln x)$$

$$\gg\gg \quad \frac{1}{1+e^{-\ln x}} \quad \gg\gg \quad \frac{1}{1+x^{-1}}$$

- Then using differentiation:

$$f_{e^X}(x) = \frac{d}{dx}\left(\frac{1}{1+x^{-1}}\right)$$

$$\gg\gg \quad -(1+x^{-1})^{-2} \cdot (-x^{-2}) \quad \gg\gg \quad \frac{1}{(x+1)^2}$$

Continuous wait times

06 Theory

Exponential variable

A random variable X is **exponential**, written $X \sim \mathrm{Exp}(\lambda)$, when X measures the *wait time until first arrival* in a Poisson process with rate λ.

Exponential PDF:

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

- The exponential distribution is the continuous counterpart of the geometric distribution.
    - Analogous to how the Poisson distribution is a like a continuous binomial.
- Notice that:

$$\int_0^\infty e^{-\lambda t}\,dt \quad \gg\gg \quad -\lambda^{-1}(e^{-\lambda\cdot\infty}-1) \quad \gg\gg \quad \lambda^{-1}$$

so the coefficient of λ in f_X is there to ensure that $P[-\infty \leq X \leq \infty] = 1$.
- Notice also that the "tail probability" $P[X > t]$ is given by $e^{-\lambda t}$, an exponential decay.
    - Compute the improper integral to find this.

Erlang variable

A random variable X is **Erlang**, written $X \sim \mathrm{Erlang}(\ell, \lambda)$, when X measures the *wait time until ℓ^{th} arrival* in a Poisson process with rate λ.

Erlang PDF:

$$f_X(t) = \frac{\lambda^\ell}{(\ell-1)!} t^{\ell-1} e^{-\lambda t}$$

- The Erlang distribution is the continuous counterpart of the Pascal distribution.

07 Illustration

Example - Earthquake wait time

Suppose the San Andreas fault produces major earthquakes modeled by a Poisson process, with an average of 1 major earthquake every 100 years.

- (a) What is the probability that there will *not* be a major earthquake in the next 20 years?
- (b) What is the probability that *three* earthquakes will strike within the next 20 years?

Solution
(a)
Since the average wait time is 100 years, we set $\lambda = 0.01$ earthquakes per year. Set $X \sim \mathrm{Exp}(0.01)$ and compute:

$$P[X > 20] = e^{-\lambda\cdot 20} \quad \gg\gg \quad e^{-0.01\cdot 20} \quad \gg\gg \quad \approx 0.82$$

(b)
The same Poisson process has the same $\lambda = 0.01$ earthquakes per year. Set

$X \sim \text{Erlang}(3, 0.01)$, so:

$$f_X(t) = \frac{\lambda^\ell}{(\ell-1)!} t^{\ell-1} e^{-\lambda t}$$

$$\gg\gg \quad \frac{(0.01)^3}{(3-1)!} t^{3-1} e^{-0.01 \cdot t} \quad \gg\gg \quad \frac{10^{-6}}{2} t^2 e^{-0.01 \cdot t}$$

and compute:

$$P[X \leq 20] = \int_0^{20} f_X(x)\, dx$$

$$\gg\gg \quad \int_0^{20} \frac{10^{-6}}{2} t^2 e^{-0.01 \cdot t}\, dt \quad \gg\gg \quad \approx 0.00115$$

08 Theory

The memoryless distribution is exponential

The exponential distribution is memoryless. This means that knowledge that an event has not yet occurred does not affect the probability of its occurring in future time intervals:

$$P[X > t + s \mid X > t] = P[X > s].$$

This is easily checked using the PDF: $e^{-\lambda(t+s)} / e^{-\lambda t} = e^{-\lambda s}$.

No other continuous distribution is memoryless. This means any other (continuous) memoryless distribution agrees in probability with the exponential distribution. The reason is that the memoryless property can be rewritten as $P[X > t + s] = P[X > t]P[X > s]$. Consider $P[X > x]$ as a function of x, and notice that this function *converts sums into products*. Only the exponential function can do this.

The geometric distribution is the discrete memoryless one.

$$P[X > n] \quad \gg\gg \quad \sum_{k=n+1}^{\infty} q^{k-1} p \quad \gg\gg \quad q^n p(1 + q + q^2 + \ldots)$$

$$\gg\gg \quad q^n \frac{p}{1-q} \quad \gg\gg \quad q^n$$

and by substituting $n + k$, we also know $P[X > n + k] = q^{n+k}$.

Then:

$$P[X = n + k \mid X > n] \quad \gg\gg \quad \frac{P[X = n + k]}{P[X > n]} \quad \gg\gg \quad \frac{q^{n+k-1} p}{q^n}$$

$$\gg\gg \quad q^{k-1} p \quad \gg\gg \quad P[X = k]$$

Extra - Inversion of decay rate factor in exponential

For constants a and λ:

$$\mathrm{Exp}(a\lambda) \sim \frac{1}{a}\mathrm{Exp}(\lambda)$$

Derivation

Let $X \sim \mathrm{Exp}(\lambda)$ and observe that $P[X > t] = e^{-\lambda t}$ (the "tail probability").

Now observe that:

$$P[a^{-1}X > t] = P[X > at] = e^{-\lambda at}$$

Let $Y \sim \mathrm{Exp}(a\lambda)$. So we see that:

$$P[a^{-1}X > t] = P[Y > t]$$

Since the tail event is complementary to the cumulative event, these two distributions have the same CDF, and therefore they are equal.

Extra - Geometric limit to exponential

Divide the waiting time into small intervals. Let $p = \frac{\lambda}{n}$ be the probability of at least one success in the time interval $[a, a + \frac{1}{n}]$ for any a. Assume these events are independent.

A random variable T_n measuring the end time of the first interval $[\frac{k-1}{n}, \frac{k}{n}]$ containing a success would have a geometric distribution with $\frac{k}{n}$ in place of k:

$$P\left[T_n = \frac{k}{n}\right] = \left(1 - \frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n}$$

By taking the sum of a geometric series, one finds:

$$P[T_n > x] = \left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor}$$

Thus $P[T_n > x] \to e^{-\lambda x}$ as $n \to \infty$.