CSCI 3022 Intro to Data Science
Expectation

random outcomes 'up to'

$F(x) = P(X \leq x)$

cdf: cumulative density function:

Opening Example:

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x < 1 \\ 0 & else \end{cases}$$

pdf

1. What is the $\boxed{\text{cdf}}$ of sales for any x?
2. Find the probability that X is less than .25?
3. X is greater than .75?
4. $P(.25 < X < .75)$?

Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment.
2. Binomial: binary outcome experiment success *count* in n tries.
3. Geometric: Total trials *until a success* of a binary outcome experiment.
4. Negative Binomial: Trials until *r* binary outcome experiment *successes*.
5. Poisson: *counting* outcomes with a fixed rate λ.

Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment.
   $f(x) = p^x(1-p)^{1-x}$
   $X = 0$ or 1
2. Binomial: binary outcome experiment success *count* in n tries.
   $f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$
   $x = 0$ or 1 or $\cdots$ or n.
3. Geometric: Total trials *until a success* of a binary outcome experiment.
   $f(x) = (1-p)^{x-1} p$
   $X = 1, 2, 3, \ldots, \infty$
4. Negative Binomial: Trials until *r* binary outcome experiment *successes*.
   $f(x) = \binom{x-1}{r-1} p^r (1-p)^{(x-r)}$
   $x = r, r+1, r+2, r+3, \cdots$
5. Poisson: *counting* outcomes with a fixed rate λ.
   $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
   $X = 0, 1, 2, 3, \ldots$

Poisson: $\frac{\text{\# of events}}{\text{unit time}}$

Last Time...: the blocks of continuous probability

1. Exponential: time-until-event of a things that happen at a rate of $\lambda \frac{events}{time}$.

λ: parameter

Exp: time until exactly 1/first event

$$\underline{f(x) = \lambda e^{-\lambda x}}; \quad x \geq 0$$

2. Uniform: all events from $[a, b]$ are equally likely:

$$f(x) = \frac{1}{b-a}; \qquad x \in [a, b]$$

For continuous distributions, we can't just add up a big list of outcomes and their probabilities. Instead, the probability of *single* outcomes is always zero. We add up *intervals*, which turns into an integral:

integral = Σ sum

$$P(\underset{\text{lower bound}}{a} \leq X \leq \underset{\text{upper bound}}{b}) = \int_a^b f(x)\, dx$$

tells us the probability of all outcomes from a to b of a continuous RV with pdf $f(x)$.

Percentiles of a Distribution

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ quantile of the distribution.

How can we express this in terms of $f(x), F(x)$?

Notation:

Visually:

Percentiles of a Distribution

Integration to recall

1) substitution rule
2) Power rule
3) IBP
by parts

Statistics of pdfs/random variables

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or .5 quantile of the distribution.

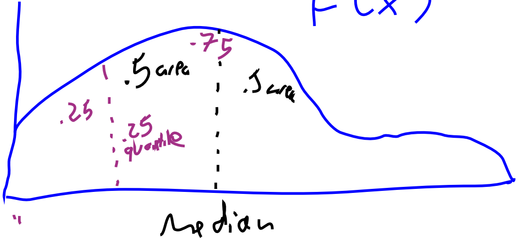
median, mean, variances, etc.

How can we express this in terms of $f(x), F(x)$?

percentiles

$f(x)$

Visually:



Notation:

$\tilde{x}$ satisfies $F(\tilde{x}) = .5$, or

$.5 = F(\tilde{x})$

$$.5 = \int_{-\infty}^{\tilde{x}} f(x)\, dx$$

First quartile: x value so that 1/4 area is left of it:

$.25 = \int_{-\infty}^{x} f(t)\,dt$: $.25 = F(x_q)$

Opening Solution

cdf: $P(X \leq x)$ a version of $P(\cancel{a \leq} X \leq b)$ no lower bound

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \boxed{\frac{3}{2}(1 - x^2)} & \boxed{0 \leq x < 1} \\ 0 & else \end{cases}$$

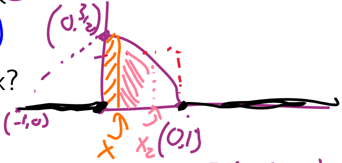
$f(x) = \frac{3}{2}(1 - x^2)$

$= \frac{3}{2}(1-x)(1+x)$

quadratic: downward

y-intercept $(x=0, \frac{3}{2})$

x-intercepts $(x=1, 0)$, $(x=-1, 0)$



1. What is the cdf of sales for any x?

$F(x)$ = area "up to x"
$= P(0 \leq X \leq x)$ start at 0 ok: $P(X<0) = 0 = \int_{-\infty}^{0} f(x)\,dx$

GOAL: $\int_0^x \frac{3}{2}(1 - y^2)\,dy$

$= \int_0^x \frac{3}{2} - \frac{3y^2}{2}\,dy$

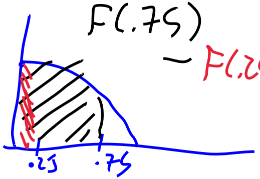
2. Find the probability that X is less than .25?
3. X is greater than .75?
4. $P(.25 < X < .75)$?

Opening Solution

$P(X > .75) = 1 - P(X \le .75) = 1 - F(.75)$

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

all prob lives from $[0,1]$



$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x < 1 \\ 0 & else \end{cases}$$

$\int \frac{3}{2}\, dy = \frac{3y}{2}$

$+ \int \frac{3y^2}{2}\, dy = \frac{3}{2} \cdot \frac{y^3}{3} = \frac{3y}{2} - \frac{y^3}{2} \Big|_0^x$

1. What is the cdf of sales for any x?
   $F(x) = P(X \le x) = \boxed{\int_0^x \frac{3}{2}(1 - t^2)\, dt}$
   $F(x) = \boxed{\frac{3x}{2} - \frac{x^3}{2}}$ works in $(0,1)$.
   if $x < 0$, $F(x) = 0$
   if $x > 1$, $F(x) = 1$.
2. Find the probability that X is less than .25? $F(.25)$
   $\frac{3(.25)}{2} - \frac{(.25)^3}{2}$
3. X is greater than .75? $1 - F(.75)$
4. $P(.25 < X < .75)$? $F(.75) - F(.25)$
   $\int_{.25}^{.75} f(x)\, dx = \boxed{F(.75)} - \boxed{F(.25)}$
   all up to .75 minus the $(0 : .25)$ area.

Pops and Samples

Today marks the start of a large jump in how we approach data science problems:

1. We know about *sample statistics* like $\bar{X}$, s_X. apply to a list of #s
2. We've defined some *processes* that gives rise to distributions like the binomial, exponential, etc.
3. **Now:** we start bridging the gap! Given data and sample statistics, how do we estimate or infer properties of the underlying reality process? (parameters like p, λ).

   To do this, we need an understanding of centrality and dispersion of a process or density function might be.

Mean/Expected Value

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

1% of students take 1 class

2% take 7 classes.

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

Students pay more money when enrolled in more courses, and so the university wants to know what the *average* number of courses students take per semester.

Mean/Expected Value

recall: $\bar{X} = \frac{\sum x_i}{n}$

\# of times we saw x-values

~ probability

Definition: *Expected Value:*
For a discrete random variable X with pdf f(x), the *expected* value or *mean* value of X is denoted as E(X) and is calculated as:

Mean/Expected Value

$$\frac{\sum X_i}{n}$$ Sum outputs/outcomes.

Definition: *Expected Value:*
For a discrete random variable X with pdf f(x), the *expected* value or *mean* value of X is denoted as E(X) and is calculated as:

Prob of each outcome

$$E[X] = \sum_{x\in\Omega} x \cdot \boxed{P(X = x)}$$

times the outcome itself

Mean/Expected Value

Example:, cont'd:

The pdf of X is given to you as follows:

multiply

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01	.03	.13	.25	.39	.17	.02

outcomes

Prob of outcome

$x \cdot f(x)$ .01, .06, .39, 1.00, 1.95, 1.02, .14

What is $E[X]$?

Mean/Expected Value

data = [1, 2, 2, 2, 3, 3, 3 ... 3, 4, 4, ... 4, ...]

13 of em   25

average = 4.57

Example:, cont'd:

The pdf of X is given to you as follows:

$\bar{X}$? mean

x = outcome	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02
count = 100 · P(x)	1	3	13	25	39	17	2

→ $\bar{X} = 4.57$

What is $E[X]$?

large impact

$$E[X] = \sum_{x \in \Omega} x \cdot P(X = x) = 1 \cdot .01 + 2 \cdot .03 + 3 \cdot .13 + 4 \cdot .25 + 5 \cdot .39 + 6 \cdot .17 + 7 \cdot .02$$

small impact

$$\frac{\sum X}{n} = \frac{1 + 2 + 2 + 2 + 3 + 1 \cdots + 4 + 4 \cdots}{n}$$

$$E[X] = 4.57$$

$= \frac{1}{n} + 3\left(\frac{2}{n}\right) + 13\left(\frac{3}{n}\right) + 25\left(\frac{4}{n}\right) + \cdots$ $n = 100$, total

$= \frac{1}{100} \cdot 1 + 3 \cdot \frac{2}{100} + 13 \cdot \frac{3}{100} = 1 \cdot \frac{1}{100} + \frac{3}{100} \cdot 2 + \frac{13}{100} \cdot 3 + \cdots$

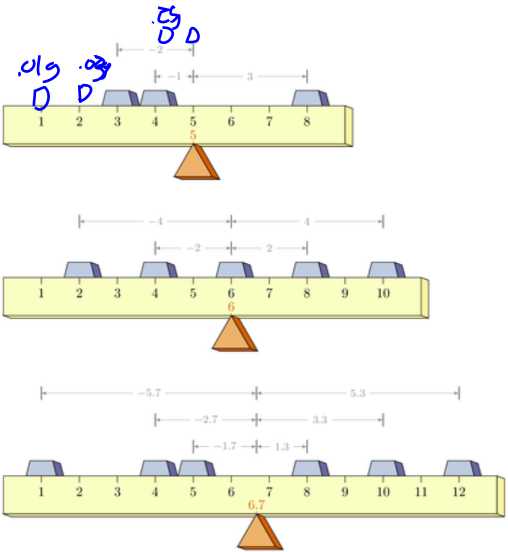
Interpreting Expected Value: Relative Frequency

One way to interpret expected value of a discrete distribution (especially on a finite support) is the sample mean if we managed to observe observations that *exactly* mirror the probability mass function.

In the preceding example, the pmf was given at 7 values of X with a precision up to 1%. In this case, if we had exactly 100 students and their proportions *observed* exactly mirrored the probabilities given in the example, the sample mean would be identical to the population mean.

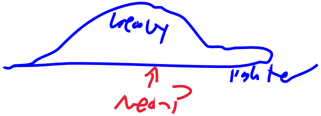
Interpreting Expected Value

x	1	2
$f(x)$:	.01	.03



- The "center of mass" of a set of point masses
- Each mass exerts an "$r \times f$" force on the balancing point.
- Same idea holds in continuous space: we're looking for a centroid of an object.

http://www.texample.net/media/tikz/examples/TEX/balance.tex



Mean/Expected Value

discrete: $E[X] = \sum x \cdot f(x)$

"expected value of" all outcomes outcome · prob of outcome

Definition: *Expected Value:*
For a continuous random variable X with pdf $f(x)$, the *expected* value or *mean* value of X is denoted as $E(X)$ and is calculated as:

$E[\ ]$
input random variables

$f(x)$
input #s.

$$E[X] = \int x \ f(x) \ dx$$

all outcomes outcome · "Prob of outcome"

Mean/Expected Value

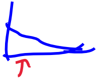
Definition: *Expected Value:*
For a continuous random variable X with pdf $f(x)$, the *expected* value or *mean* value of X is denoted as $E(X)$ and is calculated as:

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

all "outcomes" ; bounds need to hold
all non-zero regions of $f(x)$.

Mean/Expected Value

$f(x) = \lambda e^{-\lambda x}$



$x \in [0, \infty)$

Example:
The lifetime (in years) of a certain brand of battery is exponentially distributed with $\lambda = 0.25$.

mean of exponential

How long, on average, will the battery last?

$$\text{Goal: } E[X] = \int_{\text{all outcomes}} x \cdot f(x)\, dx = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_0^\infty x e^{-\lambda x}\, dx$$

recall: this polynomial x times exponential e^{-x}

Integration by parts: integral product rule.

Mean/Expected Value

$$\int \underbrace{a}_{\text{original}} \underbrace{db}_{\text{integrand}} = ab - \int b\, da$$

Example:
The lifetime (in years) of a certain brand of battery is exponentially distributed with $\lambda = 0.25$.

choose as "u"

L og
I nv. trig
$\boxed{\text{P oly}} = u$
$\boxed{\text{E xponentially}} = dv$
T rig

How long, on average, will the battery last?

$$\lambda \int \underbrace{\boxed{x}}_{u} \underbrace{\boxed{e^{-\lambda x}\, dx}}_{dv}$$

$$du = \frac{d}{dx} x = 1$$

$$\int dv = \int e^{-\lambda x}\, dx \Rightarrow v = \frac{1}{-\lambda} e^{-\lambda x}$$

Recall: Integration by Parts: $\int \underset{1}{u}\, \underset{2}{dv} = uv - \int v\, du$. Mental shortcuts: "integration product rule," $\boxed{\text{"LIPET"}}$ "ultraviolet voodoo":

Mean/Expected Value

Example:

The lifetime (in years) of a certain brand of battery is exponentially distributed with $\lambda = 0.25$.

How long, on average, will the battery last?

Start with $E[X] = \int_{\infty}^{\infty} x f(x)\, dx$, then use our known $f(x)$:

$E[X] = \int_0^\infty \lambda x e^{-\lambda x}\, dx$, now via IBP with $\boxed{u = \lambda x;}\ \boxed{dv = e^{-\lambda x}\ ;}$

$E[X] = \lambda \underset{u}{x}\left(\frac{-1}{\lambda}\underset{v}{e^{-\lambda x}}\right)\Big|_0^\infty - \int_0^\infty \underset{v}{\lambda\left(\frac{-1}{\lambda}\right.} \underset{du}{e^{-\lambda x}}\left.\right)\, dx$

Both xe^{-x} and e^{-x} $\underline{\to 0}$ as $x \underline{\to \infty}$, so we're left with:

$E[X] = \frac{-1}{\lambda} e^{-\lambda x}\big|_0^\infty$ which is just $1/\lambda$. This should come as no surprise, since we interpret λ as an average *rate* in events-per-time, but the exponential measures time-until-event, so the expected value of the exponential is the reciprocal of the rate!

Expected Value of a Function

If a discrete r.v. X has a a density $P(X = x)$, then the expected value of any function $g(X)$ is computed as:

1. Continuous:

2. Discrete:

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(x)$ is substituted in place of x.

Expected Value of a Function

If a discrete r.v. X has a a density $P(X = x)$, then the expected value of any function $g(X)$ is computed as:

1. Continuous:

$$E[X] = \int_{\infty}^{\infty} xf(x)\ dx$$

2. Discrete:

$$E[X] = \sum_{x} xf(x)\ dx$$

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(x)$ is substituted in place of x.

Expected Value of a Function

Example: A random variable X has pdf:

$$f(x) = \frac{3}{4}(1 - x^2); \quad -1 \leq X \leq 1$$

What is $E(X^3)$?

Expected Value of a Function

Example: A random variable X has pdf:

$$f(x) = \frac{3}{4}(1 - x^2); \quad -1 \leq X \leq 1$$

What is $E(X^3)$?

$$E(X^3) = \int_{-1}^{1} x^3 \frac{3}{4}(1 - x^2)\, dx = \frac{3x^4}{16} - \frac{3x^6}{24} |_{-1}^{1} = 0$$

Review: What is $F(x)$?

Expected Value of a Function

Example: A random variable X has pdf:

$$f(x) = \frac{3}{4}(1 - x^2); \quad -1 \leq X \leq 1$$

What is $E(X^3)$?

$$E(X^3) = \int_{-1}^{1} x^3 \frac{3}{4}(1 - x^2)\, dx = \frac{3x^4}{16} - \frac{3x^6}{24}|_{-1}^{1} = 0$$

Review: What is $F(x)$?

$$F(x) = \int_{-1}^{x} f(t)\, dt = \frac{3t}{4} - \frac{3t^3}{12}|_{-1}^{x}$$

Expected Value of a Linear Function

If $g(X)$ is a linear function of X (i.e., $g(X) = aX + b$) then $E[g(X)]$ can be easily computed from $E(X)$.

Theorem:
Let $a, b \in \mathbb{R}$ and X be a random variable with density f. Then:

Proof:

Note: This works for continuous and discrete random variables.

Expected Value of a Linear Function

If $g(X)$ is a linear function of X (i.e., $g(X) = aX + b$) then $E[g(X)]$ can be easily computed from $E(X)$.

Theorem:
Let $a, b \in \mathbb{R}$ and X be a random variable with density f. Then:

$$E[g(X)] = g(E[X])$$

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

Proof:
$E[aX + b] = \int (ax + b) f(x)\, dx = a \int x f(x)\, dx + b \int f(x)\, dx = aE[X] + b$, since integration is also linear!
Note: This works for continuous and discrete random variables.

Linear Expectation

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

Earlier, we calculated that $E(X)$ was 4.57. If students pay \$500 per course plus a \$100 per-semester registration fee, what is the average amount of money the university can expect a student to pay each a semester?

Linear Expectation

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

Earlier, we calculated that $E(X)$ was 4.57. If students pay $500 per course plus a $100 per-semester registration fee, what is the average amount of money the university can expect a student to pay each a semester?

$Money = 500 \cdot Courses + 100 = 500X + 100 = g(X)$. Then,

$$E[g(X)] = g(E[X]) = 500 \cdot 4.57 + 100 = 2385.$$

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

> Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that
>
> $$s = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}$$

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

$$s = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \quad \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

$$s = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

We might ask: what is the *expected* value of how spread out x-value are?

Population variance is this idea expressed as an expectation:

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

$$s = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \quad \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

Population variance is this idea expressed as an expectation:

$$Var[X] = E[\underbrace{(X - E[X])^2}_{\text{squared deviations}}] = E[(X - \mu_X)^2]$$

Expectation Practice

Final exercise: find $E[X]$ of the *exponential* distribution.

Expectation Practice

Final exercise: find $E[X]$ of the *exponential* distribution.

Daily Recap

Today we learned

1. Expectation

Moving forward:

- nb day Friday!

Next time in lecture:

- Expected dispersion/spread: calculating variances from pdfs!