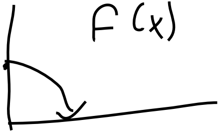
CSCI 3022-002 Intro to Data Science
Expectation

$P(a \leq X \leq b) = \int_a^b f(x)\,dx$



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

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

Start

$P(X \leq -40) = 0$

$P(X \leq 4) = 1$

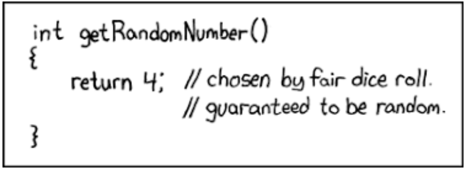
What is the cdf of sales for any x?

Recall:

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

Announcements and Reminders

- Practicum posted soon! No HW next Monday!



Last Time...: the blocks of continuous probability

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

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

2. Uniform: all events form $[a, b]$ 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:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

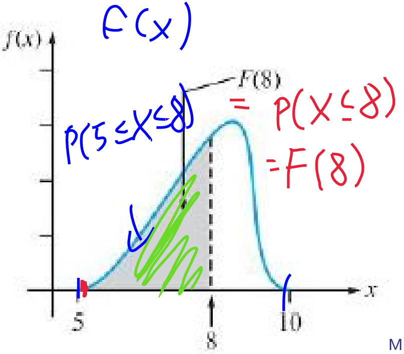
tells us the probability of all outcomes from a to b of a continuous RV with pdf $f(x)$.

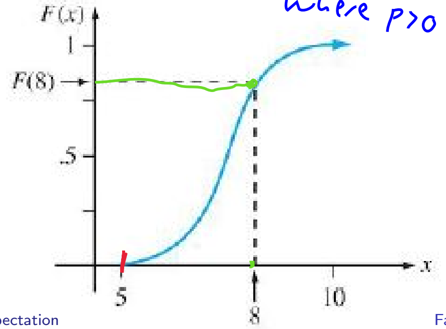
Cumulative Density Function

Definition: *Cumulative Density Function*
The *cumulative distribution function* (cdf) is denoted with $F(x)$. For a continuous r.v. X with pdf $f(x)$, $F(x)$ is defined for every real number x by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

→ or at "the start" of X where P>0





A cdf example:

$$F(x) = \frac{3x}{2} - \frac{x^3}{2} = P(X \leq x)$$

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

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

$$F(x) = \int_0^x \frac{3}{2}(1 - a^2)\, da$$

$$= \int_0^x \frac{3}{2} - \frac{3a^2}{2}\, da$$

$$= \frac{3a}{2} - \frac{\cancel{3}a^3}{2 \cdot \cancel{3}} \Big|_0^x$$

0 → not always 0!

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

   quick way to get probs:

2. Find the probability that X is less than .25?

   $$P(X < .25) = F(.25) = \left(\frac{3x}{2} - \frac{x^3}{2}\right) - 0$$

3. X is greater than .75?

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

4. $P(.25 < X < .75)$?

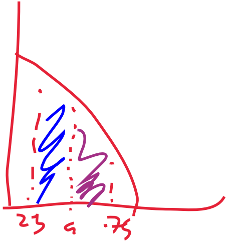
   $$\int_{.25}^{.75} f(x)\,dx$$

A cdf example:

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

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



1. What is the cdf of sales for any x?
   $F(x) = P(X \leq x) = \int_0^x \frac{3}{2}(1 - t^2)\, dt$
   $F(x) = \frac{3x}{2} - \frac{x^3}{2}$
2. Find the probability that X is less than .25? $F(.25)$
3. X is greater than .75? $1 - F(.75)$
4. $P(.25 < X < .75)$? $\boxed{F(.75) - F(.25)}$

$$\int_{.25}^{.75} f(x)\, dx$$

split at $a = 0$ a

$$\underbrace{\int_{.25}^{a=0} f(x)dx}_{= \int_0^{.25} f(x)dx = -F(.25)} + \underbrace{\int_{a=0}^{.75} f(x)dx}_{F(.75)}$$

Continuous CDFs

Wait, we've seen this before...

Recall: *The Fundamental Theorem of Calculus*.

Suppose F is an anti-derivative of f. Then:

1. $$\frac{d}{dx}\int_a^x f(t)\,dt = f(x);$$

   a.k.a.

   $$\frac{d}{dx}F(x) = f(x);$$

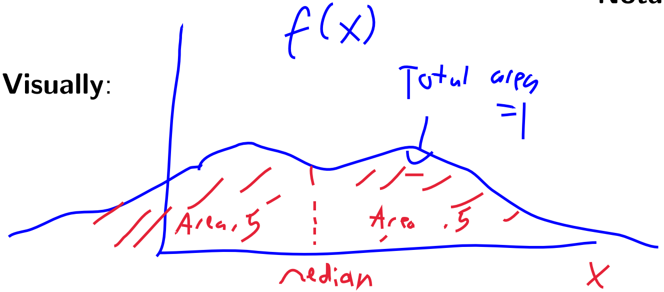
2. $$\int_a^b f(x)\,dx = F(B) - F(A).$$

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 $f(x), F(x)$?

Visually:



Notation:

$\tilde{x}$ satisfies:

$.5 =$ Area to left / Area to right

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

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

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

Visually:

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

Probability Recaps

1. **Discrete:** find probabilities in the probability mass function
$$f(x) = P(X = x).$$
2. **Continuous:** find probabilities by integrating the probability density function
$$\int_a^b f(x)\,dx = P(a < X < b).$$
3. We can find cumulative probabilities or probability on ranges of outcomes in the cumulative density function
$$F(x) = P(X \leq x) = \sum_{X \leq x} f(x) \text{ or } \int_{-\infty}^{x} f(t)\,dt$$
4. **Definition**: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ *quantile* of the distribution.
$\tilde{x}$ satisfies $F(\tilde{x}) = .5$, or
$$.5 = \int_{\infty}^{\tilde{x}} f(x)\,dx$$

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 $\boxed{\bar{X},\ s_X.}$ Mean & variance of process
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:

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

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

$P(X=2)=.5$
$P(X=4)=.5$ $\Rightarrow$ AVG = 3!

what if $P(X=2)=.75$?

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:

add up... times its probability

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

where each x outcome is

Mean/Expected Value

$E[X] = \sum_{\text{outcomes}} (\text{outcome}) \cdot P(\text{of that outcome})$

Example:, cont'd:

The pdf of X is given to you as follows:

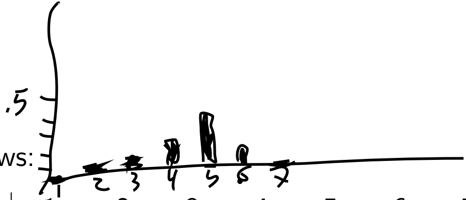
middle?

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01	.03	.13	.25	.39	.17	.02
$x \cdot P(X=x)$	.01	.06	.39	1.00	1.95	1.02	.14

What is $E[X]$?

sum these!

Mean/Expected Value



Example:, cont'd:

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

→ sum to 1 (100%)

What is $E[X]$?

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

$$E[X] = 4.57$$

Interpreting Expected Value: Relative Frequency

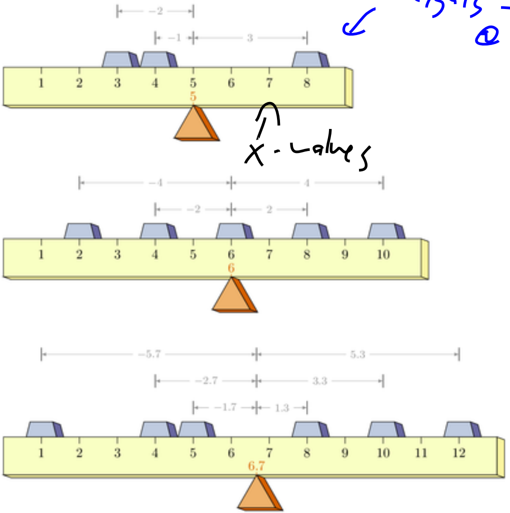
ex: exactly 100 student 25 took 4 classes
39 took 5 class...

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

weights = probability @ that x-value



x-values

- 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

Recall: discrete

$\sum x \, P(X=x)$

$E[X] = \sum_{\text{outcomes}} x \cdot f(x)$

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:

$$E[X] = \int_{\text{all outcomes}} x f(x) dx$$

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:

"all outcomes"

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

Mean/Expected Value

Recall: $f(x) = \lambda e^{-\lambda x}$ from $[0, \infty)$

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?

exp(.25)

$E[X]$ for $X \sim \exp(\lambda)$

$$\int_{\text{outcomes}} x f(x)\,dx = \int_0^\infty x\left(\lambda e^{-\lambda x}\right)dx$$

1) symbolab

Mean/Expected Value

$$E[x] = \int x \cdot f(x)\,dx = \int \underset{p}{\lambda x}\, \underset{e}{e^{-\lambda x}}\,dx$$

Example:

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

LIPET: 1) Log 2) Inv. trig 3) Poly 4) Exponential 5) Trig

How long, on average, will the battery last?

$u =$ poly: λx $\quad du = \lambda\,dx$

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

$a = -\lambda x$, $da = -\lambda\,dx$, $\frac{da}{-\lambda} = dx$ $\rightarrow \int e^{a}\,\frac{da}{-\lambda} = \frac{1}{-\lambda}\int e^{a} = \frac{1}{-\lambda} e^{-\lambda x}$

$\frac{d}{dx} e^{-\lambda x} = -\lambda e^{-\lambda x}$

Recall: Integration by Parts: $\int u\,dv = uv - \int v\,du$. Mental shortcuts: "integration product rule," "LIPET"

choose for u: $\quad$ unit.okt $\quad$ voodoo

Mean/Expected Value

$E[X] =$ "how many years on average, 'til 1 burnout?" $= 1/\lambda$: years/burnout

Example:
The lifetime (in years) of a certain brand of battery is exponentially distributed with $\lambda = 0.25$. $= \frac{\text{burnouts}}{\text{year}}$

How long, on average, will the battery last?

$u = \lambda x \quad du = \lambda\, dx$

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

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 $u = \lambda x$; $dv = e^{-\lambda x}$:
$E[X] = \lambda x (\frac{-1}{\lambda} e^{-\lambda x})|_0^{\infty} - \int_0^{\infty} \lambda (\frac{-1}{\lambda} e^{-\lambda x})\, dx$

$x e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx$

Both xe^{-x} and $e^{-x} \to 0$ as $x \to \infty$, so we're left with:
$E[X] = \frac{-1}{\lambda} e^{-\lambda x}|_0^{\infty}$ 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!

$\lim_{x \to \infty} x e^{-x} = \lim \frac{x}{e^{x}} \qquad e^{x} >> x^{(a)}$

$-x \cancel{e^{-\lambda x}} - \frac{1}{\lambda} \cancel{e^{-\lambda x}} \Big|_{x=0}^{x \to \infty} = \frac{1}{\lambda}$

$0 \qquad 0 \qquad \underbrace{}_{-(\frac{-1}{\lambda} e^{0})}$

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:

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

2. Discrete:

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

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 $E(X^3)$?

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 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,

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

Daily Recap

Today we learned

1. Expected Value

Moving forward:

- nb day Friday!

Next time in lecture:

- Population Variances