# Probability Primer

**Agenda** Discrete uniform distribution, Expected Value $\mathbb{E}$, Bernoulli, Binomial, linearity of expectation, Independence, Variance, Normal distribution, continuous probability, density, cdf, conditional probability, law of total probability, Bayes Rule, Poisson

**HW** { • watch Wyner's probability lectures from Moneyball

**Dice** Let X represent the roll of a dice,

$$X = \begin{cases} 1 & \text{with probability } 1/6 \\ 2 & \text{wp } 1/6 \\ 3 & \\ 4 & \\ 5 & \dots \quad 1/6 \\ 6 & \end{cases}$$

Formally, $\boxed{X \sim \text{Uniform}(\{1,2,3,4,5,6\})}$.

X is a discrete random variable

discrete because the set of outcomes $\Omega = \{1,2,3,4,5,6\}$ is finite (countable).

Q What is the average value of a dice roll?

$$\frac{1+2+3+4+5+6}{6} = 3.5$$

But why is 3.5 the average value?

One nice interpretation is this, if you roll a dice n= 10 trillion times, you'd expect

just about 10 trillion 1's, 10 trillion 2's, etc.
and the numeric average of these values will be just about 3.5 (the frequency argument).

* What if now we define a new dice $Y$ as the Random variable

$$Y = \begin{cases} 1 & \text{wp } 0 \\ 2 & \text{wp } 2/6 \\ 3, 4, 5, 6 & \text{wp } 1/6 \end{cases}$$

What then is the value of an average dice Roll?
Again consider the frequency argument.
If we Roll $n = 60$ trillion dice, then
10 trillion each will be 3, 4, 5, 6
20 trillion will be 2
So the sum of all the dice Rolls is

$$\frac{2(20 \text{ trillion}) + 3(10 \text{ trillion}) + 4(10 \text{ trillion}) + 5(10 \text{ trillion}) + 6(10 \text{ trillion})}{20 \text{ trillion} + 10 \text{ trillion} + \cdots + 10 \text{ trillion}}$$

$$= \frac{2 \cdot 2 + 3 + 4 + 5 + 6}{6} = \frac{22}{6} = 3.67$$

* Written in another way,

$$2\left(\frac{2}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

$$= \frac{2 \cdot 2 + 3 + 4 + 5 + 6}{6} = 3.67$$

So, simply multiply each outcome by its probability and then sum!

**Def** The expected value of a discrete random variable $X$ is

$$\mathbb{E}X = \sum_{x} x \, \mathbb{P}(X = x)$$

- discrete random variable $X$ means $X$ can be one of countably many outcomes $\Omega$
  e.g. Dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$
- The summation $\sum_{x}$ is over the outcomes $x \in \Omega$
- $\mathbb{P}(X = x)$ is the probability that the random variable $X$ equals the value $x$
  e.g. Dice, $x = 5$, $\mathbb{P}(X = 5) = 1/6$

$\boxed{\text{Coin Flip}}$ Let $X$ represent a coin flip,

$$X = \begin{cases} 1 & \text{with probability } p \quad \text{(heads)} \\ 0 & \text{w.p. } 1-p \quad \text{(tails)} \end{cases}$$

Formally, $\boxed{X \sim \text{Bernoulli}(p)}$

$p$ here is a <u>parameter</u>, $p \in [0,1]$.

The expected value of a Bernoulli($p$) random variable, or informally the average value of a biased coin flip, is

$$\begin{aligned} \mathbb{E}X &= \sum_x x \, \mathbb{P}(X = x) \\ &= \sum_{x \in \{0,1\}} x \, \mathbb{P}(X = x) \\ &= 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) \\ &= \mathbb{P}(X=1) \\ &= p. \end{aligned}$$

# Shaqtin' a fool

A free throw shooter named ________
takes n free throw shots.
He makes a free throw with prob. p.
The random variable X representing the number of shots he makes is

$$X = \sum_{i=1}^{n} X_i$$

$$X_i \overset{iid}{\sim} \text{Bernoulli}(p)$$

iid = independent and identically distributed

i.e., $X_i = \mathbb{1}\left\{\begin{matrix} i\text{th shot} \\ \text{is made} \end{matrix}\right\} = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$

X is the sum of n Bernoulli(p) R.V.'s,

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

How many shots will he make on average?
Well, the expected value of a Binomial R.V. is

$$\mathbb{E}X = \sum_x x\,\mathbb{P}(X=x) = \sum_{x=0}^{n} x\,\mathbb{P}(X=x)$$

and $\mathbb{P}(X=x) = \mathbb{P}(\text{make } x \text{ of } n \text{ shots}, \quad x \in \{0,\dots,n\})$

for one ordering, e.g. $\underbrace{\underline{1}\ \underline{1}\ \underline{1} \cdots \underline{1}}_{x \text{ made shots}}\ \underbrace{\underline{0}\ \underline{0} \cdots \underline{0}}_{\substack{n-x \\ \text{missed shots}}}$

the probability that this exact ordering occurs is

$$\underbrace{p \cdot p \cdot p \cdot \cdots p}_{x \text{ made shots}} \cdot \underbrace{(1-p)(1-p)\cdots(1-p)}_{\substack{n-x \text{ made} \\ \text{shots}}}$$

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

and the number of such orderings with $x$ 1's is

$$\binom{n}{x} = n \text{ choose } x = \frac{n!}{x!\,(n-x)!}$$

so $\mathbb{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

so $\mathbb{E}X = \sum_x x\,\mathbb{P}(X=x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^x$

How the fuck do we sum this though?

The **Binomial Theorem** goes through this, using induction. But there's an easier way!

**Theorem (Linearity of Expectation)**

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

and $\mathbb{E}[cX] = c\,\mathbb{E}[X]$, $c \in \mathbb{R}$

Idea Average value of the sum of 2 dice rolls is the sum of the averages! Think about it, using the frequency argument!

**Proof Idea**

$$\mathbb{E}[X+Y] = \sum_x \sum_y (x+y)\, \mathbb{P}(X=x, Y=y)$$

↳ and

$$= \sum_x x \sum_y \mathbb{P}(X=x, Y=y) + \sum_y y \sum_x \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x\, \mathbb{P}(X=x) + \sum_y y\, \mathbb{P}(Y=y)$$

$$= \mathbb{E}X + \mathbb{E}Y.$$

and $\mathbb{E}[cX] = \sum_{cx} cx\, \mathbb{P}(cX = cx)$

$$= \sum_{x} cx\, \mathbb{P}(X = x)$$

$$= c \sum_{x} x\, \mathbb{P}(X = x)$$

$$= c\, \mathbb{E}X.$$

Therefore, when $X \sim \text{Binomial}(n, p)$,

$X = \sum_{i=1}^{n} X_i$ where $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$,

$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}X_i = \sum_{i=1}^{n} p = np.$ Easy!

So, if ____ takes $n$ shots, then he will on average make $np$ of them!

Now, recall for $X \sim \text{Binomial}(n, p)$, the number of made shots, that $X = \sum_{i=1}^{n} X_i$ where $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$.

<u>iid</u> means independent and identically distributed.

identically distributed here means each $X_i$ is similarly a

Bernoulli$(p)$ random variable.

But what does independence mean formally?

**def** $X$ and $Y$ are <u>independent</u> random variables means

$$\mathbb{P}(X=x,\ Y=y) = \mathbb{P}(X=x)\cdot\mathbb{P}(Y=y)$$

* Why is this definition a good one?

Frequency argument:

1 trillion pairs of independent shots.

$\approx p$ trillion made 1st shots

- of these, $\approx p$ fraction are paired with a 2nd shot make
- of these, $\approx 1-p$ fraction are paired with a 2nd shot miss

$\approx 1-p$ trillion missed 1st shots

- $\approx p$
- $\approx 1-p$

$\Downarrow$

$\approx p^2$ trillion pairs (made, made)

$\approx p\cdot(1-p)$ trillion pairs (made, missed)

$\approx (1-p)\cdot p$ trillion pairs (missed, made)

$\approx (1-p)^2$ trillion pairs (missed, missed) $\square$

**Thm** If $X$ and $Y$ are independent then

$$\mathbb{E}(XY) = (\mathbb{E}X) \cdot (\mathbb{E}Y)$$

**Proof** Assume $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)$.

Then
$$\begin{aligned}
\mathbb{E}(X \cdot Y) &= \sum_x \sum_y xy \, \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y xy \, \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y) \\
&= \sum_x x \, \mathbb{P}(X=x) \cdot \sum_y y \, \mathbb{P}(Y=y) \\
&= (\mathbb{E}X) \cdot (\mathbb{E}Y).
\end{aligned}$$

Recall if $X \sim \text{Binomial}(n,p)$ then $\mathbb{E}X = np$,

so ______ will make $np$ free throws on average.

But what will the spread or deviation from this spread look like ??

**Def** The variance of a random variable $X$ is the expected squared deviation from its mean,

$$\text{VAR}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

Large variance = on average, $X$ is far from its mean, since $(X-\mathbb{E}X)^2$ is large often.

**Theorem (Funny Variance Formula)**

$$\boxed{\mathrm{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2}$$

Proof

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}(X-\mathbb{E}X)^2 \\
&= \mathbb{E}\left(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2\right) \\
&= \mathbb{E}X^2 - \mathbb{E}[2X\mathbb{E}X] + \mathbb{E}\left((\mathbb{E}X)^2\right)
\end{aligned}$$

by linearity of expectation

$$= \mathbb{E}X^2 - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2$$

by linearity of expectation because $\mathbb{E}X \in \mathbb{R}$, $(\mathbb{E}X)^2 \in \mathbb{R}$ are real numbers

$$= \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

So, the variance of ______'s number of made free throws is

$$\mathrm{var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$
$$= \mathbb{E}X^2 - (np)^2$$

and $\mathbb{E}X^2 = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2$ $\qquad X_i \overset{iid}{\sim} \mathrm{Ber}(p)$

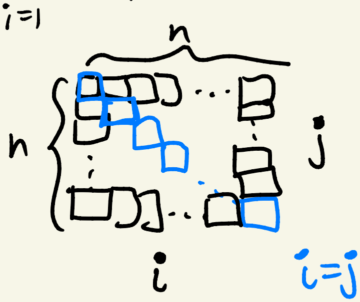
$$= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right)$$
$$= \mathbb{E}\left(\sum_{i \neq j} X_i X_j + \sum_{i=1}^{n} X_i\right)$$
$$= \sum_{i \neq j} \mathbb{E}(X_i X_j) + \sum_{i=1}^{n} \mathbb{E}X_i$$

by linearity of expectation

$$= \sum_{i \neq j} (\mathbb{E}X_i)\cdot(\mathbb{E}X_j) + \sum_{i=1}^{n} \mathbb{E}X_i$$

by independence

$$= \sum_{i \neq j} p \cdot p + \sum_{i=1}^{n} p$$
$$= (n^2 - n)\cdot p^2 + n \cdot p$$




Then $\mathrm{var}(X) = n^2p^2 - np^2 + np - n^2p^2 = n(p - p^2)$
$$= np(1-p).$$

The standard deviation of a R.V. $X$ is the square root of the variance,

$$sd(X) = \sqrt{VAR(X)}$$

Because $VAR(X) = \mathbb{E}(X - \mathbb{E}X)^2$ is on the squared scale, the s.d. uses a more grounded unit of measurement.

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

$$X = \sum_{i=1}^{n} X_i, \quad X_i \overset{iid}{\sim} \text{Ber}(p)$$

$$\mathbb{E}X = np, \quad sd(X) = \sqrt{np(1-p)}$$

Central limit theorem the (Rescaled) sum of iid random variables $X_i$ with mean $\mu = \mathbb{E}X_i$ and s.d. $\sigma = \sqrt{VAR(X_i)}$, $S_n = \sum_{i=1}^{n} X_i$, converges to a standard Normal (Gaussian) Random Variable $Z \sim \mathcal{N}(0,1)$,

$$\lim_{n \to \infty} \mathbb{P}\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \mathbb{P}(a \le Z \le b)$$

$X \sim \text{Binomial}(n, p)$ is $S_n = X$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx N(0,1) \quad \text{for } n \text{ large}$$

$$S_n \approx N(n\mu, \sqrt{n}\,\sigma)$$

$$S_n = X, \quad \mu = p, \quad \sigma = \sqrt{p(1-p)}$$

So,
$$X \approx N\left(np, \sqrt{np(1-p)}\right).$$

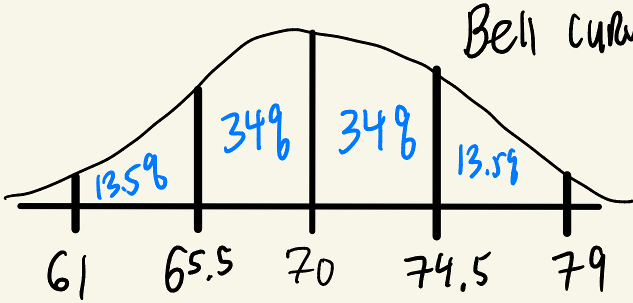
______ takes $n = 100$ shots.

Assume we know his $p = .7$

$$\sqrt{np(1-p)} = 10\sqrt{(.7)(.3)} = 10\sqrt{.21} = 10(.458) \approx 4.5$$

$$np = 100(.7) = 70$$

Bell curve!




Normal dist area values.

* But what Really <u>is</u> the normal distribution?

Discrete probability: $\mathbb{P}(X = x)$ defines a R.V.

Now, if $X \sim N(\mu, \sigma^2)$, $X$ can take on any Real number in $\mathbb{R}$, so if $P(X=x) > 0$ for each $x \in \mathbb{R}$ then $\sum_{x \in \mathbb{R}} \mathbb{P}(X=x) = \infty \neq 1$, a contradiction...

So, in <u>continuous probability</u>, where a Random variable $X$ can take on any of uncountably many values in $\mathbb{R}$, $X$ is defined by its

$$\boxed{\underline{\text{density function}}\ f_X(x) = \mathbb{P}(X \in [x, x+dx))}$$

I hope you like calculus!

Let's start with the <u>CDF (cumulative distribution function)</u>

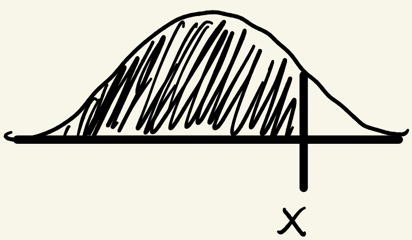
$$F_X(x) = \mathbb{P}(X \leq x)$$

which is easy to understand.

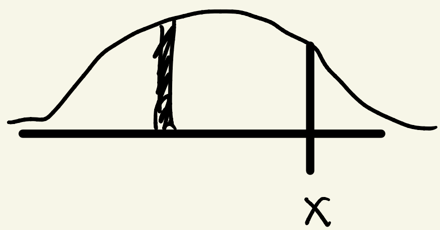
Well, $F(x) = \mathbb{P}(X \le x)$

$$= \int_{-\infty}^{x} \mathbb{P}(X \in [x, x+dx))\, dx$$

by def. of integral as area under curve







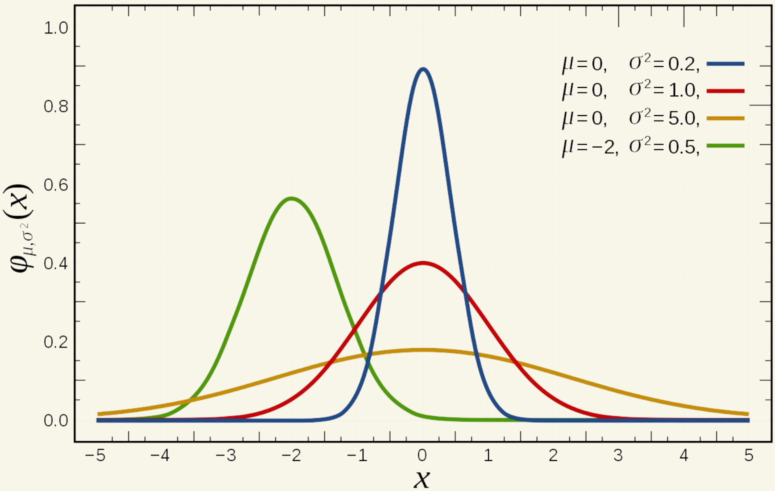


$$= \int_{-\infty}^{x} f_X(x)\, dx$$

So we define the density as the derivative of the cdf, $\boxed{f_X(x) = F_X'(x)}$

**Def** The normal distribution with mean $\mu$ and variance $\sigma^2$, written $X \sim N(\mu, \sigma^2)$,

has density $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$




Before, in discrete probability, $\mathbb{E}X = \sum_x x\, \mathbb{P}(X=x)$

Now, in continuous probability, $\mathbb{P}(X=x)$ does not exist, and is replaced by $f(x) = \mathbb{P}(X \in [x, x+dx))$.

The analog of $\sum_x$ in continuous space is $\int_x$ because we sum up infinitesimally small rectangles.

**Def** The <u>expectation</u> of a continuous R.V. $X$ is

$$\mathbb{E}X = \int x f(x)\,dx$$

and of a function of $X$ is

$$\mathbb{E}g(X) = \int g(x) f(x)\,dx.$$

If $X \sim N(\mu, \sigma^2)$ then

$$\mathbb{E}X = \int x f(x)\,dx \overset{HW}{=} \mu \quad \text{and}$$

$$\mathbb{E}X^2 = \int x^2 f(x)\,dx \overset{HW}{=} \sigma^2 \Longrightarrow \mathrm{Var}(X) = \sigma^2.$$

The <u>standard normal distribution</u> $Z \sim N(0,1)$ has density $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

and std. normal CDF $\Phi(x) := \int_{-\infty}^{x} \phi(x)\,dx$

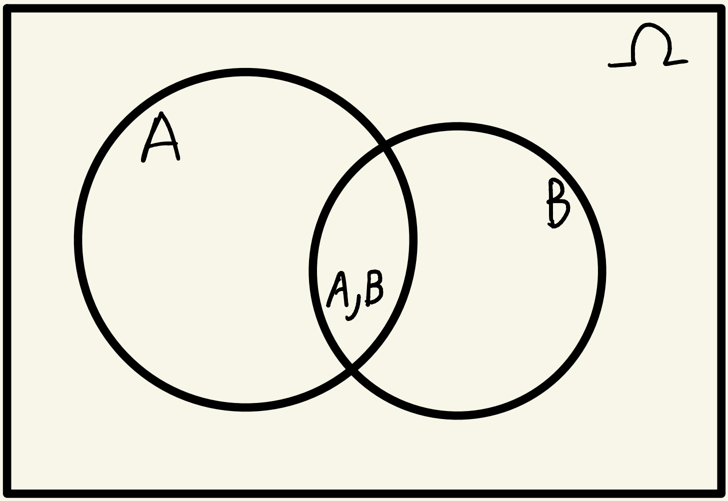
which does not have a closed form.

**def** the conditional probability of $A|B$ "A given B" is defined by

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)}$$

$A,B$ "A and B"

To understand why this is a good definition, see Venn diagram:



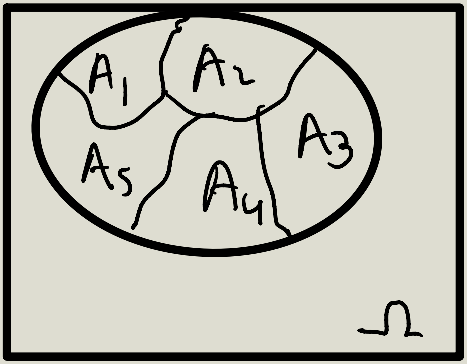

$\Omega$ = entire sample space (set of outcomes)

Given we're in B, the total prob. is $\mathbb{P}(B)$, and the prob. of hitting A is $\mathbb{P}(A,B)$, so $\mathbb{P}(A|B) := \mathbb{P}(A,B)/\mathbb{P}(B)$.

## Law of Disjoint Unions

If $A_1, \ldots, A_n$ partitions the set $A$,

so $A_1 \cup A_2 \cup \cdots \cup A_n = A$ and $A_i \cap A_j = \emptyset \quad \forall i \neq j$




then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

**Thm** If $X, Y$ are independent then $\mathbb{P}(X|Y) = \mathbb{P}(X)$

**Pf** $$\mathbb{P}(X|Y) = \frac{\mathbb{P}(X,Y)}{\mathbb{P}(Y)} = \frac{\mathbb{P}(X)\cdot\mathbb{P}(Y)}{\mathbb{P}(Y)} = \mathbb{P}(X).$$

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A|B)}{\mathbb{P}(B)}$$

and so we also have $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(B,A)}{\mathbb{P}(A)}$

$$\implies \mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B,A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

$$\boxed{\underline{\text{Bayes' Rule}} \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}}$$

* Extremely important!

Ex Suppose half of USA is male and half is female.
But 1/3 of NBA fans are female.
Also, 1/3 of Americans are NBA fans.
What's the probability of a male/female being an NBA fan?

$P(M) = \frac{1}{2}$, $P(F) = \frac{1}{2}$, $P(M/NBA) = \frac{2}{3}$, $P(F/NBA) = \frac{1}{3}$.

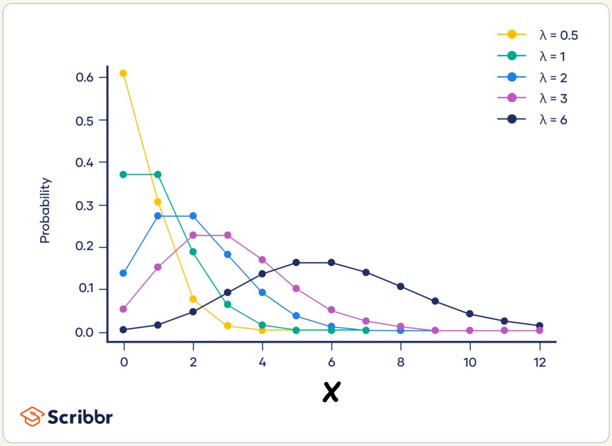
$$P(NBA/M) = \frac{P(M/NBA)\,P(NBA)}{P(M)} = \frac{\frac{2}{3}\cdot\frac{1}{3}}{\frac{1}{2}} = \frac{4}{9}$$

$P(NBA/F) = \frac{2}{9}$.

* Finally, we'll be using the Poisson distribution in this class

**def** A Poisson($\lambda$) Random Variable with $\lambda > 0$ can take on any nonnegative integer value $x = 0, 1, 2, \dots$ and has probability mass function

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$




$$\mathbb{E}X = \sum_x x\, \mathbb{P}(X = x)$$

$$= \sum_{x=0}^{\infty} x\, e^{-\lambda} \frac{\lambda^x}{x!}$$

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

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= \lambda e^{-\lambda} e^{\lambda} \quad \text{by def of Taylor series of } e^{\lambda}$$

$$= \lambda.$$

Similarly, $\mathrm{var}[X] = \lambda$.

There are a few other "canonical" distributions you should know

- Geometric
- Hypergeometric
- Exponential
- Beta
- Gamma

<u>HW</u> Watch Professor Wyner's probability lectures from Moneyball (e.g., expected value)