Statistical Foundations of Data Science

Discrete Distributions

University of the Witwatersrand

2022

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx$

Lesson Plan

- Review Question

Lesson Plan

- Review Question
- The notion of a random variable

Lesson Plan

- Review Question
- The notion of a random variable
- Means and Variances

Lesson Plan

- Review Question
- The notion of a random variable
- Means and Variances
- Some standard distributions

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx =$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx =$
- $\frac{1}{2} x^2 e^{-2x}]_0^\infty + \int_0^\infty x e^{-2x} dx =$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx =$
- $\frac{1}{2} x^2 e^{-2x}]_0^\infty + \int_0^\infty x e^{-2x} dx =$
- $\frac{-1}{2} x e^{-2x}]_0^\infty + \frac{1}{2} \int_0^\infty e^{-2x} dx =$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx =$
- $\frac{1}{2} x^2 e^{-2x}]_0^\infty + \int_0^\infty x e^{-2x} dx =$
- $\frac{-1}{2} x e^{-2x}]_0^\infty + \frac{1}{2} \int_0^\infty e^{-2x} dx =$
- $\frac{1}{2} \int_0^\infty e^{-2x} dx =$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx =$
- $\frac{1}{2} x^2 e^{-2x}]_0^\infty + \int_0^\infty x e^{-2x} dx =$
- $\frac{-1}{2} x e^{-2x}]_0^\infty + \frac{1}{2} \int_0^\infty e^{-2x} dx =$
- $\frac{1}{2} \int_0^\infty e^{-2x} dx =$
- $\frac{1}{4}$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx$
- $\int_0^\infty \frac{y^2}{4} e^{-y} \frac{dy}{2}$ for $y = 2x$

Review Question

- Compute the intergral $\int_0^\infty x^2 e^{-2x} dx$
- $\int_0^\infty \frac{y^2}{4} e^{-y} \frac{dy}{2}$ for $y = 2x$
- $\frac{1}{8}\Gamma(3) = \frac{1}{4}$

Random variables

- There is a formal measure-theoretic definition, which we'll ignore.

Random variables

- There is a formal measure-theoretic definition, which we'll ignore.
- For now it can be thought of as an unknown quantity.

Random variables

- There is a formal measure-theoretic definition, which we'll ignore.
- For now it can be thought of as an unknown quantity.
- Defined on some sample space.

Random variables

- There is a formal measure-theoretic definition, which we'll ignore.
- For now it can be thought of as an unknown quantity.
- Defined on some sample space.
- Example. We roll a standard 6 sided dice.

Random variables

- There is a formal measure-theoretic definition, which we'll ignore.
- For now it can be thought of as an unknown quantity.
- Defined on some sample space.
- Example. We roll a standard 6 sided dice.
- Random variables have well defined distributions. In the dice rolling example.

$$p(X) = \begin{cases} \frac{1}{6} & x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & elsewhere \end{cases}$$

Properties of a distribution

- Probabilities are non-negative real numbers which "sum" to 1.

Properties of a distribution

- Probabilities are non-negative real numbers which "sum" to 1.
- This week we'll look at discrete distributions. Which is to say distributions that have point masses.

Properties of a distribution

- Probabilities are non-negative real numbers which "sum" to 1.
- This week we'll look at discrete distributions. Which is to say distributions that have point masses.
- Next week we'll look at continuous distributions which have density functions.

Properties of a distribution

- Probabilities are non-negative real numbers which "sum" to 1.
- This week we'll look at discrete distributions. Which is to say distributions that have point masses.
- Next week we'll look at continuous distributions which have density functions.
- Mixed distributions also exist but are rarely studied.

Expected value

- The first property of a distribution is it's expected value. Informally this may be thought of as an average.

Expected value

- The first property of a distribution is it's expected value. Informally this may be thought of as an average.
- $\mathbb{E}[X] = \sum_x xP(X = x) = \sum_s X(s)P(s)$

Expected value

- The first property of a distribution is it's expected value. Informally this may be thought of as an average.
- $\mathbb{E}[X] = \sum_x xP(X = x) = \sum_s X(s)P(s)$
- Let's compute the expected value for our dice roll example.

Expected value

- The first property of a distribution is it's expected value. Informally this may be thought of as an average.
- $\mathbb{E}[X] = \sum_x xP(X = x) = \sum_s X(s)P(s)$
- Let's compute the expected value for our dice roll example.
- $\mathbb{E}[X] = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

Linearity fo Expectation

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Linearity fo Expectation

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
-

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_s (X+Y)(s)P(s) \\
&= \sum_s (X(s)+Y(s))P(s) \\
&= \sum_s X(s)P(s) + Y(s)P(s) \\
&= \sum_s X(s)P(s) + \sum_s Y(s)P(s) \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Dice example!

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Dice example!
- $\mathbb{E}[X] = 3.5$

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Dice example!
- $\mathbb{E}[X] = 3.5$
- If $Y = X - \mathbb{E}[X]$

Variance

- The second property of a distribution is it's variance. Informally this is usually thought of as a measure of spread.
- $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- Dice example!
- $\mathbb{E}[X] = 3.5$
- If $Y = X - \mathbb{E}[X]$
- $$p(Y) = \begin{cases} \frac{1}{6} & x \in \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\} \\ 0 & elsewhere \end{cases}$$

Variance

- Set $Z = Y^2$ then

Variance

- Set $Z = Y^2$ then
- $$p(Z) = \begin{cases} \frac{1}{3} & x \in \{0.25, 2.25, 6.25\} \\ 0 & elsewhere \end{cases}$$

Variance

- Set $Z = Y^2$ then
- $$p(Z) = \begin{cases} \frac{1}{3} & x \in \{0.25, 2.25, 6.25\} \\ 0 & elsewhere \end{cases}$$
- $\mathbb{E}[Z] = V(X) = \frac{0.25+2.25+6.25}{3} = \frac{8.75}{3} = \frac{35}{12}$

Variance - Computational Form

- $V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Variance - Computational Form

- $V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
-

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

Sanity check

- $\mathbb{E}[X^2] = \frac{1}{6}1 + \frac{1}{6}4 + \frac{1}{6}9 + \frac{1}{6}16 + \frac{1}{6}25 + \frac{1}{6}36 = 91/6$

Sanity check

- $\mathbb{E}[X^2] = \frac{1}{6}1 + \frac{1}{6}4 + \frac{1}{6}9 + \frac{1}{6}16 + \frac{1}{6}25 + \frac{1}{6}36 = 91/6$
- $\mathbb{E}[X] = 3.5$

Sanity check

- $\mathbb{E}[X^2] = \frac{1}{6}1 + \frac{1}{6}4 + \frac{1}{6}9 + \frac{1}{6}16 + \frac{1}{6}25 + \frac{1}{6}36 = 91/6$
- $\mathbb{E}[X] = 3.5$
- $V(X) = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$

Sanity check

- $\mathbb{E}[X^2] = \frac{1}{6}1 + \frac{1}{6}4 + \frac{1}{6}9 + \frac{1}{6}16 + \frac{1}{6}25 + \frac{1}{6}36 = 91/6$
- $\mathbb{E}[X] = 3.5$
- $V(X) = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$
- So formula seems to work

Discrete Uniform Distribution

- $$p(X) = \begin{cases} \frac{1}{m} & x \in \{1, 2, ..., m\} \\ 0 & elsewhere \end{cases}$$

Discrete Uniform Distribution

- $$p(X) = \begin{cases} \frac{1}{m} & x \in \{1, 2, ..., m\} \\ 0 & elsewhere \end{cases}$$
- Dice is an example

Discrete Uniform Distribution

- $$p(X) = \begin{cases} \frac{1}{m} & x \in \{1, 2, ..., m\} \\ 0 & elsewhere \end{cases}$$
- Dice is an example
- $\mathbb{E}[X] = \frac{m+1}{2}$

Discrete Uniform Distribution

- $$p(X) = \begin{cases} \frac{1}{m} & x \in \{1, 2, ..., m\} \\ 0 & elsewhere \end{cases}$$
- Dice is an example
- $\mathbb{E}[X] = \frac{m+1}{2}$
- $V(X) = \frac{m^2-1}{12}$

Discrete Uniform Distribution

- $$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{m} i\frac{1}{m} \\ &= \frac{1}{m}\sum_{i=1}^{m} i \\ &= \frac{1}{m}\frac{m(m+1)}{2} \\ &= \frac{m+1}{2}\end{aligned}$$

Discrete Uniform Distribution

- $$\begin{aligned}\mathbb{E}[X^2] &= \sum_{i=1}^{m} i^2 \frac{1}{m} \\ &= \frac{1}{m}\sum_{i=1}^{m} i^2 \\ &= \frac{1}{m}\frac{m(m+1)(2m+1)}{6} \\ &= \frac{(m+1)(2m+1)}{6}\end{aligned}$$

Discrete Uniform Distribution

- $$\begin{aligned}\mathbb{E}[X^2] &= \sum_{i=1}^{m} i^2 \frac{1}{m} \\ &= \frac{1}{m}\sum_{i=1}^{m} i^2 \\ &= \frac{1}{m}\frac{m(m+1)(2m+1)}{6} \\ &= \frac{(m+1)(2m+1)}{6}\end{aligned}$$
- $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(m+1)(2m+1)}{6} - \frac{(m+1)^2}{4} = \frac{m^2-1}{12}$

Bernoulli and Binomial

- Bernoulli is a coin flip.

Bernoulli and Binomial

- Bernoulli is a coin flip.
- $$p(X) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & elsewhere \end{cases}$$

Bernoulli and Binomial

- Bernoulli is a coin flip.
- $$p(X) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & elsewhere \end{cases}$$
- $\mathbb{E}[X] = p$

Bernoulli and Binomial

- Bernoulli is a coin flip.
- $$p(X) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & elsewhere \end{cases}$$
- $\mathbb{E}[X] = p$
- $\mathbb{E}[X^2] = p$ so $V(X) = p - p^2 = p(1 - p)$

Bernoulli and Binomial

- Bernoulli is a coin flip.
- $$p(X) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & elsewhere \end{cases}$$
- $\mathbb{E}[X] = p$
- $\mathbb{E}[X^2] = p$ so $V(X) = p - p^2 = p(1 - p)$
- Binomial is what happens when you generate n independent Bernoulli distributions and add them.

Bernoulli and Binomial

- Bernoulli is a coin flip.
- $$p(X) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & elsewhere \end{cases}$$
- $\mathbb{E}[X] = p$
- $\mathbb{E}[X^2] = p$ so $V(X) = p - p^2 = p(1 - p)$
- Binomial is what happens when you generate n independent Bernoulli distributions and add them.
- $$p(X) = \begin{cases} \binom{n}{k} p^k (1 - p)^{N-k} & x \in \{0, 1, 2, ..., n\} \\ 0 & elsewhere \end{cases}$$

Binomial

- $\mathbb{E}[X] = np$ by linearity of expectation

Binomial

- $\mathbb{E}[X] = np$ by linearity of expectation
- $V(X) = np(1-p)$, variances are linear FOR INDEPENDENT RANDOM VARIABLES

Binomial

- $\mathbb{E}[X] = np$ by linearity of expectation
- $V(X) = np(1-p)$, variances are linear FOR INDEPENDENT RANDOM VARIABLES
- $$p(X) = \begin{cases} \binom{n}{k} p^k (1-p)^{N-k} & x \in \{0, 1, 2, ..., n\} \\ 0 & \textit{elsewhere} \end{cases}$$

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.
- $$p(X) = \begin{cases} (1-p)^k p & x \in \{0, 1, 2, ..., n\} \\ 0 & elsewhere \end{cases}$$

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.
- $$p(X) = \begin{cases} (1-p)^k p & x \in \{0, 1, 2, \ldots, n\} \\ 0 & \mathit{elsewhere} \end{cases}$$
- Is a distribution. $\Sigma_{i=0}^{n}(1-p)^i p = 1$

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.
- $$p(X) = \begin{cases} (1-p)^k p & x \in \{0, 1, 2, \ldots, n\} \\ 0 & \textit{elsewhere} \end{cases}$$
- Is a distribution. $\Sigma_{i=0}^{n}(1-p)^i p = 1$
- Mean $\Sigma_{i=0}^{n} i(1-p)^i p$

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.
- $$p(X) = \begin{cases} (1-p)^k p & x \in \{0, 1, 2, ..., n\} \\ 0 & \textit{elsewhere} \end{cases}$$
- Is a distribution. $\Sigma_{i=0}^{n}(1-p)^i p = 1$
- Mean $\Sigma_{i=0}^{n} i(1-p)^i p$
- $$\begin{aligned} M &= 1(1-p)p + 2(1-p)^2 p + 3(1-p)^3 p.. \\ (1-p)M &= 1(1-p)^2 p + 2(1-p)^3 p + 3(1-p)^4 p.. \\ pM &= (1-p)p + (1-p)^2 p + (1-p)^3 p + \\ M &= \frac{1-p}{p} \end{aligned}$$

Geometric Distribution

- Flip a (possibly) biased coin (probability of heads p).
- X counts the number of tails before getting a head.
- $$p(X) = \begin{cases} (1-p)^k p & x \in \{0, 1, 2, ..., n\} \\ 0 & elsewhere \end{cases}$$
- Is a distribution. $\Sigma_{i=0}^{n}(1-p)^i p = 1$
- Mean $\Sigma_{i=0}^{n} i(1-p)^i p$
- $$\begin{aligned} M &= 1(1-p)p + 2(1-p)^2 p + 3(1-p)^3 p.. \\ (1-p)M &= 1(1-p)^2 p + 2(1-p)^3 p + 3(1-p)^4 p.. \\ pM &= (1-p)p + (1-p)^2 p + (1-p)^3 p + \\ M &= \frac{1-p}{p} \end{aligned}$$
- Variance $\frac{1-p}{p^2}$

Poisson Distribution

- Originally developed for counting events in a time period.

Poisson Distribution

- Originally developed for counting events in a time period.
- It can be thought of as a limit of the binomial distribution as $n \to \infty$ and np held constant. Here p is decreasing.

Poisson Distribution

- Originally developed for counting events in a time period.
- It can be thought of as a limit of the binomial distribution as $n \to \infty$ and np held constant. Here p is decreasing.
- You are waiting for an event to happen for example receiving an e-mail.

Poisson Distribution

- Originally developed for counting events in a time period.
- It can be thought of as a limit of the binomial distribution as $n \to \infty$ and np held constant. Here p is decreasing.
- You are waiting for an event to happen for example receiving an e-mail.
- Every hour you receive one or you don't.

Poisson Distribution

- Originally developed for counting events in a time period.
- It can be thought of as a limit of the binomial distribution as $n \to \infty$ and np held constant. Here p is decreasing.
- You are waiting for an event to happen for example receiving an e-mail.
- Every hour you receive one or you don't.
- Well sometimes you receive more than one in an hour. Every minute you either receive one or you don't?

Poisson Distribution

- Originally developed for counting events in a time period.
- It can be thought of as a limit of the binomial distribution as $n \to \infty$ and np held constant. Here p is decreasing.
- You are waiting for an event to happen for example receiving an e-mail.
- Every hour you receive one or you don't.
- Well sometimes you receive more than one in an hour. Every minute you either receive one or you don't?
- Every second or milli-second or micro-second. The point is that as we take the limit that the odds of getting two at once goes to zero.

Poisson Distribution

- Within a particular time period there is therefore a rate $\lambda (= np)$ of an event occurring.

Poisson Distribution

- Within a particular time period there is therefore a rate $\lambda (= np)$ of an event occurring.
- The sum of two (or more) poisson variables is again poisson with rate parameters added.

Poisson Distribution

- Within a particular time period there is therefore a rate $\lambda (= np)$ of an event occurring.
- The sum of two (or more) poisson variables is again poisson with rate parameters added.
- For example if you're counting the number of events that occur between noon and 2pm that's the number of events between noon and 1pm plus the number of events between 1pm and 2pm.

Poisson Distribution

$$
\begin{aligned}
\mathbb{P}(X = k) &= \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \binom{n}{k} \frac{\lambda}{n}^k (1 - \frac{\lambda}{n})^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \frac{\lambda^k}{n^k} e^{-\lambda} \\
&= \lim_{n\to\infty} \frac{n(n-1)(n-2)..(n-k+1)}{k!} \frac{\lambda^k}{n^k} e^{-\lambda} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \frac{\lambda^k}{k!} e^{-\lambda}
\end{aligned}
$$

Poisson Distribution - a distribution?

$$
\begin{aligned}
\sum_{k=0}^{\infty} \mathbb{P}(X = k)] &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} e^{\lambda} \\
&= 1
\end{aligned}
$$

Poisson Distribution - mean

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\
&= \lambda \sum_{t=0}^{\infty} \frac{\lambda^t}{(t)!} e^{-\lambda} \\
&= \lambda
\end{aligned}$$

Poisson Distribution - variance

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda}$$
$$= \sum_{k=0}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda}$$

This is ugly to deal with.

Poisson Distribution - variance

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} \\
&= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}e^{-\lambda} \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} \\
&= \lambda^2
\end{aligned}
$$

Poisson Distribution - variance

$$
\begin{aligned}
V(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}
$$

Negative binomial Distrbution

- Geometric we wait for the 1st head.

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!
- That is $\frac{r(1-p)}{p}$ and $\frac{r(1-p)}{p^2}$

Negative binomial Distrbution

- Geometric we wait for the 1st head.

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!
- That is $\frac{r(1-p)}{p}$ and $\frac{r(1-p)}{p^2}$

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!
- That is $\frac{r(1-p)}{p}$ and $\frac{r(1-p)}{p^2}$
- What about $\mathbb{P}(X = k)$?

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!
- That is $\frac{r(1-p)}{p}$ and $\frac{r(1-p)}{p^2}$
- What about $\mathbb{P}(X = k)$?
- Well that is in the first $k + r - 1$ clips we get k tails and $r - 1$ heads followed by a head on flip $k + r$.

Negative binomial Distrbution

- Geometric we wait for the 1st head.
- Negative binomial is the waiting time for the r^{th} head.
- Mean and variance are trivial to compute. They're r times the mean and variance for a geometic!
- That is $\frac{r(1-p)}{p}$ and $\frac{r(1-p)}{p^2}$
- What about $\mathbb{P}(X = k)$?
- Well that is in the first $k + r - 1$ clips we get k tails and $r - 1$ heads followed by a head on flip $k + r$.
- $\mathbb{P}(X = k) = \binom{k+r-1}{k}(1-p)^k p^r$

Problems

A beginner archer calculates that he has hit the target less than 90 percent of the time (over his entire career of shooting up to that point). After some more practice sessions he gets better and eventually recalculates his (again lifetime) hit rate. He is pleased to notice that he's now hit the target more than 90 percent of the time. Is he guaranteed to have hit it at exactly 90 percent of the time at some point between the two times he calculated it? Give a proof that the average must be exactly 90 percent at some point or provide a counter-example.

Solution

Consider the function $f(n) = h - 0.9n$ where n is the number of shots and h is the number of hits by time n. Not that at each time time step f either increases by 0.1 or decreases by 0.9 and that f is always an integer multiple of 0.1. It must therefore hit the value 0 in order to cross it from the negative side.

Problems

Monty Hall decides to shake up his game show. He now has seven doors behind two of which are cars, one car is red and the other is blue and behind the other five doors are goats. The doors are labelled A,B,C, D, E, F and G. Our protagonist chooses a door. Monty will then reveal two goats and offer the player a chance to switch. Compute:

1. Find the probability of winning either car if the player does not switch. [2]
2. Find the probability of winning either car if the player switches. [3]
3. Find the probability of winning the red car if we does not switch. [2]
4. Find the probability of winning the red car the player switches. [3]

Solution

1. Find the probability of winning either car if the player does not switch. [2]
   $\frac{2}{7}$
2. Find the probability of winning either car if the player switches. [3]
   $\frac{2}{7} \cdot \frac{1}{4} + \frac{5}{7} \cdot \frac{2}{4} = \frac{3}{7}$
3. Find the probability of winning the red car if we does not switch. [2]
   $\frac{1}{7}$
4. Find the probability of winning the red car the player switches. [3]
   $\frac{1}{7} \cdot \frac{0}{4} + \frac{1}{7} \cdot \frac{1}{4} + \frac{5}{7} \cdot \frac{1}{4} = \frac{3}{14}$

Problem

A darts player shoots at an unmarked target, that is she either hits or the target or does not. She hits on her first throw and misses on her second. Thereafter the probability that she hits the next shot is equal to the proportion of shots she has hit so far. Compute with proof the probability that she hits exactly 50 of her first 100 shots?

Solution

$\frac{1}{99}$.

It turns out that the number of hits after n shots is uniform between 1 and $n-1$. Note that this is true for $n=2$ and proceed by induction on n. Assume true for $n=k$.

$$\begin{aligned}
\mathbb{P}(X_{k+1}=r) &= \frac{k-r}{k}\mathbb{P}(X_k=r) + \frac{r}{k}\mathbb{P}(X_k=r-1) \\
&= \frac{k-r}{k}\frac{1}{k-1} + \frac{r-1}{k}\frac{1}{k-1} \\
&= \frac{1}{k} \\
&= \frac{1}{(k+1)-1}
\end{aligned}$$

Probably worth doing the edge cases to but thye work out the same