Miscellaneous discrete random variable examples

Math 30530, Fall 2013

September 27, 2013

Negative Binomial

10% of the population has bloodtype AB-. A researcher samples from a pool of volunteers until he has found 5 people with bloodtype AB-. How likely is it to take 45 attempts?

Negative Binomial

10% of the population has bloodtype AB-. A researcher samples from a pool of volunteers until he has found 5 people with bloodtype AB-. How likely is it to take 45 attempts?

45th attempt has to be success; of the first 44 attempts, exactly 4 must be success, so

$$p = \binom{44}{4}(.1)^5(.9)^{40} \approx .02$$

Negative Binomial

10% of the population has bloodtype AB-. A researcher samples from a pool of volunteers until he has found 5 people with bloodtype AB-. How likely is it to take 45 attempts?

45th attempt has to be success; of the first 44 attempts, exactly 4 must be success, so

$$p = \binom{44}{4}(.1)^5(.9)^{40} \approx .02$$

Negative binomial random variable: Parameters m and p

- Counts number of independent trials (each with success probability p) until exactly m successes have been recorded
- Possible values: $k = m, m+1, m+2, \ldots$
- Mass function: $\Pr(X = k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}$

Negative Binomial

10% of the population has bloodtype AB-. A researcher samples from a pool of volunteers until he has found 5 people with bloodtype AB-. How likely is it to take 45 attempts?

45th attempt has to be success; of the first 44 attempts, exactly 4 must be success, so

$$p = \binom{44}{4}(.1)^5(.9)^{40} \approx .02$$

Negative binomial random variable: Parameters m and p

- Counts number of independent trials (each with success probability p) until exactly m successes have been recorded
- Possible values: $k = m, m+1, m+2, \ldots$
- Mass function: $\Pr(X = k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}$

When $m = 1$, Negative binomial becomes geometric

Hypergeometric

The math department has 20 faculty members, among whom 7 are women. We choose an undergraduate committee by selecting 8 faculty members at random (all choices of 8 people equally likely). What is the probability that exactly 4 of the committee members are women?

Hypergeometric

The math department has 20 faculty members, among whom 7 are women. We choose an undergraduate committee by selecting 8 faculty members at random (all choices of 8 people equally likely). What is the probability that exactly 4 of the committee members are women?

Sampling *without* replacement, so can't use binomial. Instead count number of successful outcomes, divided by total number of outcomes.

$$p = \frac{\binom{7}{4}\binom{13}{4}}{\binom{20}{8}} \approx .1986$$

(compare $\Pr(\text{Binomial}(8, .35) = 4) \approx .1875$)

Hypergeometric, continued

Hypergeometric random variable: Parameters N, M and r

- Sample, without replacement, r things from a set that has N "good" things and M "bad" things, and count the number of "good" things in the sample
- Possible values: $k = 0, 1, \ldots, r$
- Mass function: $\Pr(X = k) = \frac{\binom{N}{k}\binom{M}{r-k}}{\binom{N+M}{r}}$

Hypergeometric, continued

Hypergeometric random variable: Parameters N, M and r

- Sample, without replacement, r things from a set that has N "good" things and M "bad" things, and count the number of "good" things in the sample
- Possible values: $k = 0, 1, \ldots, r$
- Mass function: $\Pr(X = k) = \frac{\binom{N}{k}\binom{M}{r-k}}{\binom{N+M}{r}}$

When N, M large, negative binomial very close to binomial with $p = r/M$

Who wins the series?

Red Sox and Rays play a best-of-$(2n-1)$ series (first to n wins series). Red Sox win each game with probability p, all games independent.

Who wins the series?

Red Sox and Rays play a best-of-$(2n-1)$ series (first to n wins series).
Red Sox win each game with probability p, all games independent.

Q1: Let X be length of series. What's the mass function of X?

Who wins the series?

Red Sox and Rays play a best-of-$(2n-1)$ series (first to n wins series). Red Sox win each game with probability p, all games independent.

Q1: Let X be length of series. What's the mass function of X?

A: Possible values $n, \ldots, 2n-1$.

$$\begin{aligned}
\Pr(X = k) &= \binom{k-1}{n-1} p^n (1-p)^{k-n} + \binom{k-1}{n-1} (1-p)^n (1-p)^{k-n} \\
&= (\text{Red Sox win}) + (\text{Rays win})
\end{aligned}$$

Who wins the series?

Red Sox and Rays play a best-of-$(2n-1)$ series (first to n wins series). Red Sox win each game with probability p, all games independent.

Q1: Let X be length of series. What's the mass function of X?

A: Possible values $n, \ldots, 2n-1$.

$$\begin{aligned}\Pr(X=k) &= \binom{k-1}{n-1} p^n (1-p)^{k-n} + \binom{k-1}{n-1} (1-p)^n (1-p)^{k-n} \\ &= \text{(Red Sox win)} + \text{(Rays win)}\end{aligned}$$

Q2: What's the probability that the Red Sox win?

Who wins the series?

Red Sox and Rays play a best-of-$(2n-1)$ series (first to n wins series). Red Sox win each game with probability p, all games independent.

Q1: Let X be length of series. What's the mass function of X?

A: Possible values $n, \ldots, 2n-1$.

$$\begin{aligned} \Pr(X = k) &= \binom{k-1}{n-1} p^n (1-p)^{k-n} + \binom{k-1}{n-1} (1-p)^n (1-p)^{k-n} \\ &= \text{(Red Sox win)} + \text{(Rays win)} \end{aligned}$$

Q2: What's the probability that the Red Sox win?

A: $\sum_{k=n}^{2n-1} \binom{k-1}{n-1} p^n (1-p)^{k-n}$

Who wins the series, continued

Q3: Suppose Red Sox are stronger than the Rays ($p \geq 1/2$). Do the Red Sox prefer a short or a long series?

Who wins the series, continued

Q3: Suppose Red Sox are stronger than the Rays ($p \geq 1/2$). Do the Red Sox prefer a short or a long series?

A: First look at best-of-1 versus best-of-3:

$$p \text{ versus } p^2 + 2p^2(1-p) = 3p^2 - 2p^3$$

Second is larger when $p \geq 1/2$

Who wins the series, continued

Q3: Suppose Red Sox are stronger than the Rays ($p \geq 1/2$). Do the Red Sox prefer a short or a long series?

A: First look at best-of-1 versus best-of-3:

$$p \text{ versus } p^2 + 2p^2(1-p) = 3p^2 - 2p^3$$

Second is larger when $p \geq 1/2$

Next look at best-of-3 versus best-of-5:

$$3p^2 - 2p^3 \text{ versus } p^3 + 3p^3(1-p) + 6p^3(1-p)^2$$

Second is larger when $p \geq 1/2$

Who wins the series, continued

Q3: Suppose Red Sox are stronger than the Rays ($p \geq 1/2$). Do the Red Sox prefer a short or a long series?

A: First look at best-of-1 versus best-of-3:

$$p \text{ versus } p^2 + 2p^2(1-p) = 3p^2 - 2p^3$$

Second is larger when $p \geq 1/2$

Next look at best-of-3 versus best-of-5:

$$3p^2 - 2p^3 \text{ versus } p^3 + 3p^3(1-p) + 6p^3(1-p)^2$$

Second is larger when $p \geq 1/2$

General pattern: the longer series is better for the stronger team; see homework solutions for a complete verification

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Let X be the number. Possible values are $0, 1, 2, \ldots, 100$

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Let X be the number. Possible values are $0, 1, 2, \ldots, 100$

Sample space: strings of R's and L's, EITHER last is R, there are 101 R's, no more than 100 L's, OR last is L, there are 101 L's, no more than 100 R's. Probability of sample point of length n is $(1/2)^n$

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Let X be the number. Possible values are $0, 1, 2, \ldots, 100$

Sample space: strings of R's and L's, EITHER last is R, there are 101 R's, no more than 100 L's, OR last is L, there are 101 L's, no more than 100 R's. Probability of sample point of length n is $(1/2)^n$

Sample points leading to $X = k$: EITHER last is R, there are 101 R's, exactly $100 - k$ L's, OR last is L, there are 101 L's, exactly $100 - k$ R's. All have probability $(1/2)^{201-k}$.

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Let X be the number. Possible values are $0, 1, 2, \ldots, 100$

Sample space: strings of R's and L's, EITHER last is R, there are 101 R's, no more than 100 L's, OR last is L, there are 101 L's, no more than 100 R's. Probability of sample point of length n is $(1/2)^n$

Sample points leading to $X = k$: EITHER last is R, there are 101 R's, exactly $100 - k$ L's, OR last is L, there are 101 L's, exactly $100 - k$ R's. All have probability $(1/2)^{201-k}$.

$$\Pr(X = k) = 2\binom{200-k}{100}\left(\frac{1}{2}\right)^{201-k} = \frac{\binom{200-k}{k}}{2^{200-k}}$$

Banach's matchbox problem

Banach has a box of 100 matches in each pocket (left and right). Each time he lights a cigarette, he picks a random pocket to get a match from. At some moment, he reaches into a pocket and finds an empty matchbox. How many matches are in the other pocket at this moment?

Let X be the number. Possible values are $0, 1, 2, \ldots, 100$

Sample space: strings of R's and L's, EITHER last is R, there are 101 R's, no more than 100 L's, OR last is L, there are 101 L's, no more than 100 R's. Probability of sample point of length n is $(1/2)^n$

Sample points leading to $X = k$: EITHER last is R, there are 101 R's, exactly $100 - k$ L's, OR last is L, there are 101 L's, exactly $100 - k$ R's. All have probability $(1/2)^{201-k}$.

$$\Pr(X = k) = 2\binom{200-k}{100}\left(\frac{1}{2}\right)^{201-k} = \frac{\binom{200-k}{k}}{2^{200-k}}$$

See http://www-stat.stanford.edu/~susan/surprise/Banach.html