STA286 Lecture 14

Neil Montgomery

Last edited: 2017-02-26 16:24

the Binomial(n, p) distributions

Model for the number of 1s after n trials of a Bernoulli(p) process.

the Binomial(n, p) distributions

Model for the number of 1s after n trials of a Bernoulli(p) process.

pmf: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

the Binomial(n, p) distributions

Model for the number of 1s after n trials of a Bernoulli(p) process.

pmf: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

Path to $E(X)$ is easiest through the mgf:

$$M_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} = (q + pe^t)^n$$

the Binomial(n, p) distributions

Model for the number of 1s after n trials of a Bernoulli(p) process.

pmf: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

Path to $E(X)$ is easiest through the mgf:

$$M_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} = (q + pe^t)^n$$

Theorem: Let X be the sum of n independent Bernoulli(p) random variables $X_1, \ldots, X_n$. Then $X \sim$ Binomial(n, p).

the Binomial(n, p) distributions

Model for the number of 1s after n trials of a Bernoulli(p) process.

pmf: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

Path to $E(X)$ is easiest through the mgf:

$$M_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} = (q + pe^t)^n$$

Theorem: Let X be the sum of n independent Bernoulli(p) random variables $X_1, \ldots, X_n$. Then $X \sim$ Binomial(n, p).

Proof:

$$M_{X_1 + \cdots + X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t) = (q + pe^t)^n$$

the Binomial(n, p) distributions

So:

$$E(X) = E(X_1) + \cdots + E(X_n) = np$$

the Binomial(n, p) distributions

So:

$$E(X) = E(X_1) + \cdots + E(X_n) = np$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p)$$

$$SD(X) = \sqrt{np(1-p)}$$

number of Bernoulli trials until the first 1—Geometric(p)

We've done most of the work on this distribution.

$$p(y) = (1-p)^{y-1}p, \quad y \in \{1, 2, 3, \ldots\}$$

number of Bernoulli trials until the first 1—Geometric(p)

We've done most of the work on this distribution.

$$p(y) = (1 - p)^{y-1} p, \quad y \in \{1, 2, 3, \dots\}$$

$$E(Y) = \frac{1}{p}$$

number of Bernoulli trials until the first 1—Geometric(p)

We've done most of the work on this distribution.

$$p(y) = (1-p)^{y-1}p, \quad y \in \{1, 2, 3, \ldots\}$$

$$E(Y) = \frac{1}{p}$$

$$M_Y(t) = \frac{pe^t}{1 - qe^t}$$

number of Bernoulli trials until the first 1—Geometric(p)

We've done most of the work on this distribution.

$$p(y) = (1-p)^{y-1}p, \quad y \in \{1, 2, 3, \ldots\}$$

$$E(Y) = \frac{1}{p}$$

$$M_Y(t) = \frac{pe^t}{1-qe^t}$$

Variance is inevitably tedious. $E(Y^2) = \frac{d^2}{dt^2} M_Y(t)\Big|_{t=0} = \frac{2-p}{p^2}$, resulting in $\text{Var}(Y) = \frac{q}{p^2}$

number of Bernoulli trials until the first 1—Geometric(p)

We've done most of the work on this distribution.

$$p(y) = (1-p)^{y-1}p, \quad y \in \{1, 2, 3, \dots\}$$

$$E(Y) = \frac{1}{p}$$

$$M_Y(t) = \frac{pe^t}{1 - qe^t}$$

Variance is inevitably tedious. $E(Y^2) = \frac{d^2}{dt^2} M_Y(t)\Big|_{t=0} = \frac{2-p}{p^2}$, resulting in $\mathrm{Var}(Y) = \frac{q}{p^2}$

We say $Y \sim$ Geometric(p), due to the geometric rate of decay in the pmf.

cdf and the “reliability function” for a Geometric(p)

The cdf for $Y \sim \text{Geometric}(p)$ comes in handy sometimes.

$$F_Y(y) = \begin{cases} 0 & : y < 1 \\ 1 - (1-p)^k & : k \leqslant y < k+1 \end{cases}$$

cdf and the "reliability function" for a Geometric(p)

The cdf for $Y \sim$ Geometric(p) comes in handy sometimes.

$$F_Y(y) = \begin{cases} 0 & : y < 1 \\ 1 - (1-p)^k & : k \leqslant y < k+1 \end{cases}$$

The *reliability* or *survival* function of a random variable is defined as

$$R(x) = P(X > x) = 1 - P(X \leqslant x) = 1 - F(x).$$

(In class I said "left continuous"; in fact it is right continuous.)

For Y in this case (corrected from class):

$$R_Y(y) = \begin{cases} 1 & : y < 1 \\ (1-p)^k & : k \leqslant y < k+1 \end{cases}$$

the memorylessess of a Geometric(p) distribution

Bernoulli process is the discrete model for what I will (vaguely) refer to as "complete randomness", which will come to embody these (vague) ideas:

- independence (of trials)

the memorylessess of a Geometric(p) distribution

Bernoulli process is the discrete model for what I will (vaguely) refer to as "complete randomness", which will come to embody these (vague) ideas:

- independence (of trials)
- homogeneity (p doesn't change)

the memorylessess of a Geometric(p) distribution

Bernoulli process is the discrete model for what I will (vaguely) refer to as "complete randomness", which will come to embody these (vague) ideas:

- independence (of trials)
- homogeneity (p doesn't change)
- memoryless (information does not increase with number of trials)

the memorylessess of a Geometric(p) distribution

Bernoulli process is the discrete model for what I will (vaguely) refer to as "complete randomness", which will come to embody these (vague) ideas:

- independence (of trials)
- homogeneity (p doesn't change)
- memoryless (information does not increase with number of trials)

the memorylessess of a Geometric(p) distribution

Bernoulli process is the discrete model for what I will (vaguely) refer to as "complete randomness", which will come to embody these (vague) ideas:

- independence (of trials)
- homogeneity (p doesn't change)
- memoryless (information does not increase with number of trials)

The memorylessness of the Bernoulli process can be seen through a property of the Geometric(p) distribution. Suppose $Y \sim$ Geometric(p).

$$
\begin{aligned}
P(Y > j + k | Y > j) &= \frac{P(Y > j + k, Y > j)}{P(Y > j)} \\
&= \frac{P(Y > j + k)}{P(Y > j)} = \frac{(1-p)^{j+k}}{(1-p)^{j}} = (1-p)^{k} = P(Y > k)
\end{aligned}
$$

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

e.g. consider $r = 4$ and $W = 11$. $P(W = 11)$ can be derived as follows:

Trial outcomes	Probability

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

e.g. consider $r = 4$ and $W = 11$. $P(W = 11)$ can be derived as follows:

Trial outcomes	Probability
1 1 0 0 0 0 0 1 0 0 1	$(1-p)^7 p^3 p$

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

e.g. consider $r = 4$ and $W = 11$. $P(W = 11)$ can be derived as follows:

Trial outcomes	Probability
1 1 0 0 0 0 0 1 0 0 1	$(1-p)^7 p^3 p$
1 0 1 0 0 1 0 0 0 0 1	$(1-p)^7 p^3 p$

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

e.g. consider $r = 4$ and $W = 11$. $P(W = 11)$ can be derived as follows:

Trial outcomes	Probability
1 1 0 0 0 0 0 1 0 0 1	$(1-p)^7 p^3 p$
1 0 1 0 0 1 0 0 0 0 1	$(1-p)^7 p^3 p$
$\vdots$	$\vdots$

It will always be 3 1s out of 10 trials, follow by a 1. So the probability is:

$$P(W = 11) = \binom{11-1}{4-1}(1-p)^{11-4}p^4$$

number of Bernoulli trials until the r^{th} 1

Denote by W the number of trials in a Bernoulli(p) process until the r^{th} 1 occurs.

We'll say W has a "Negative Binomial" distribution, or $W \sim \text{NegBin}(r, p)$.

e.g. consider $r = 4$ and $W = 11$. $P(W = 11)$ can be derived as follows:

Trial outcomes	Probability
1 1 0 0 0 0 0 1 0 0 1	$(1-p)^7 p^3 p$
1 0 1 0 0 1 0 0 0 0 1	$(1-p)^7 p^3 p$
$\vdots$	$\vdots$

It will always be 3 1s out of 10 trials, follow by a 1. So the probability is:

$$P(W = 11) = \binom{11-1}{4-1}(1-p)^{11-4}p^4$$

In general: $P(W = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k \in \{r, r+1, r+2, \ldots\}$

NegBin(r, p)

The name comes from this version of the Binomial theorem with negative exponent:

$$(1-a)^{-r} = \sum_{k=r}^{\infty} \binom{k-1}{r-1} a^{k-r}$$

which can be used to verify that the pmf is legit.

NegBin(r, p)

The name comes from this version of the Binomial theorem with negative exponent:

$$(1-a)^{-r} = \sum_{k=r}^{\infty} \binom{k-1}{r-1} a^{k-r}$$

which can be used to verify that the pmf is legit.

Next up—mgf.

$$M_W(t) = \sum_{k=r}^{\infty} e^{tk} \binom{k-1}{r-1} (1-p)^{k-r} p^r$$

NegBin(r, p)

The name comes from this version of the Binomial theorem with negative exponent:

$$(1-a)^{-r} = \sum_{k=r}^{\infty} \binom{k-1}{r-1} a^{k-r}$$

which can be used to verify that the pmf is legit.

Next up—mgf.

$$\begin{aligned} M_W(t) &= \sum_{k=r}^{\infty} e^{tk} \binom{k-1}{r-1} (1-p)^{k-r} p^r \\ &= (pe^t)^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (qe^t)^{k-r} \\ &= \left(\frac{pe^t}{1-qe^t} \right)^r \end{aligned}$$

NegBin(r, p)

Theorem: Let W be the sum of r independent random variables $Y_1, \ldots, Y_r$ with Geometric(p) distributions. Then $W \sim$ NegBin(r, p).

Proof: $M_{Y_1+\cdots+Y_r}(t) = M_{Y_1}(t) \cdots M_{Y_r}(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$

NegBin(r, p)

Theorem: Let W be the sum of r independent random variables $Y_1, \dots, Y_r$ with Geometric(p) distributions. Then $W \sim$ NegBin(r, p).

Proof: $M_{Y_1+\cdots+Y_r}(t) = M_{Y_1}(t) \cdots M_{Y_r}(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$

So:

$$E(W) = E(Y_1) + \cdots + E(Y_r) = \frac{r}{p}$$

NegBin(r, p)

Theorem: Let W be the sum of r independent random variables $Y_1, \dots, Y_r$ with Geometric(p) distributions. Then $W \sim \text{NegBin}(r, p)$.

Proof: $M_{Y_1+\cdots+Y_r}(t) = M_{Y_1}(t) \cdots M_{Y_r}(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$

So:

$$E(W) = E(Y_1) + \cdots + E(Y_r) = \frac{r}{p}$$

and

$$\text{Var}(W) = \text{Var}(Y_1) + \cdots + \text{Var}(Y_r) = \frac{rq}{p^2}$$

the hypergeometric distributions

Reminder: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the number of ways to choose k items out of n, without replcement.

the hypergeometric distributions

Reminder: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the number of ways to choose k items out of n, without replcement.

Quality control methods often deal with this sort of situation:

100 items are produced in which 5 are defective. A sample of 10 is selected. Denote by X the number of defective items out of the 10 selected.

the hypergeometric distributions

Reminder: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the number of ways to choose k items out of n, without replcement.

Quality control methods often deal with this sort of situation:

100 items are produced in which 5 are defective. A sample of 10 is selected. Denote by X the number of defective items out of the 10 selected.

X could take on integer values between 0 and 5 inclusive, with probabilities:

$$P(X = x) = \frac{\binom{5}{x}\binom{95}{10-x}}{\binom{100}{10}} = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

N is the population size. n is the sample size. k is the number of "defective" (in this example.)

and we say X has a Hypergeometric distribution with parameters N, n, and k.

the hypergeometric distributions

I'm not sure where the name comes from. There is no natural relationship to the Bernoulli process, but there is a tangential one. . .

the hypergeometric distributions

I'm not sure where the name comes from. There is no natural relationship to the Bernoulli process, but there is a tangential one...

Hypergeometric is a "sampling without replacement" version of the Binomial. If N and n are both large, it turns out:

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\left(\frac{k}{N}\right)^x\left(1-\frac{k}{N}\right)^{n-x}$$

the hypergeometric distributions

I'm not sure where the name comes from. There is no natural relationship to the Bernoulli process, but there is a tangential one...

Hypergeometric is a "sampling without replacement" version of the Binomial. If N and n are both large, it turns out:

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\left(\frac{k}{N}\right)^{x}\left(1-\frac{k}{N}\right)^{n-x}$$

All calculations involving these distributions are tedious. It can be shown (no easy way):

$$E(X) = n\frac{k}{N} \qquad \mathrm{Var}(X) = n\frac{k}{N}\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$