Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT19

Föreläsning 6

Nancy Abdallah

Chalmers - Göteborgs Universitet

April 16, 2019

Binomial distribution

1. n repetitions of a Bernoulli trial (n is a fixed number). (The outcomes of each Bernoulli trial is either ”success” or “failure”.)
2. The trials are identical and independent, i.e. the probability of success is always the same and is denoted by p.
3. The random variable X is the number of success obtained in the n trials. X takes all the vaules from 0 to n.

The density function is given by

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

for $x = 0, \ldots, n$.

Negative Binomial distribution

1. Unfixed number of repetitions of a Bernoulli trial.
2. The trials are identical and independent.
3. The random variable X is the number of trials needed in order to get r successes, (r is a given number)

The density function of a negative binomial distribution is

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

for $r = 1, 2, 3, \ldots$ and $x = r, r+1, r+2, \ldots,$ and 0 otherwise. p and r are called the parameter of the distribution.

- If X has a negative binomial distribution with parameters r and p, then $E[X] = \frac{r}{p}$ and $Var[X] = \frac{rq}{p^2}$.

Negative Binomial distribution

1. Unfixed number of repetitions of a Bernoulli trial.
2. The trials are identical and independent.
3. The random variable X is the number of trials needed in order to get r successes, (r is a given number)

The density function of a negative binomial distribution is

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

for $r = 1, 2, 3, \ldots$ and $x = r, r+1, r+2, \ldots$, and 0 otherwise.
p and r are called the parameter of the distribution.

- If X has a negative binomial distribution with parameters r and p, then $E[X] = \frac{r}{p}$ and $Var[X] = \frac{rq}{p^2}$.

Example

An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil.

1. What is the probability that the first strike comes on the third well drilled?
2. What is the probability that the third strike comes on the seventh well drilled?
3. What is the mean and variance of the number of wells that must be drilled if the oil company wants to set up three producing wells?

Example (Solution)

1. Negative binomial distribution with parameters $r = 1$ and $p = 0.2$.
   $P(X = 3) = \binom{3-1}{1-1}(0.8)^{3-1}0.2 = 0.128$.
   Note that when $r = 1$ negative binomial distribution and geometric distribution coincide.
2. $r = 3$, $P(X = 7) = \binom{7-1}{3-1}(0.8)^{7-3}0.2^3 = 0.049$.
3. $r = 3$, $E[X] = \frac{3}{0.2} = 15$ and $Var[X] = \frac{3(0.8)}{0.2^2} = 60$.

Hypergeometric distribution

1. Draw a random sample of size n **without replacement** from a collection of N objects.
2. Among the N objects there are r that has a certain property P which we consider as "success".
3. The random variable X is the number of objects that has the property P (i.e. the number of successes).

P.S. The difference between the hypergeometric distribution and the binomial distribution is that the hypergeometric is a repetition of a binomial trial without replacement (so the trials are not identical and idenpendent).

- The density function of a hypergeometric distribution is given by

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

  for $\max[0, n-(N-r)] \leq x \leq \min(n, r)$, and 0 otherwise. N, n and r are called the parameters of the distribution.
- If X has a hypergeometric distribution then $E[X] = \frac{nr}{N}$ and $Var[X] = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = \frac{nr(N-r)(N-n)}{N^2(N-1)}$.

Example (Exercise 58 p.92)

Production line workers assemble 15 automobiles per hour. During a given hour, four are produced with improperly fitted doors. Three automobiles are selected at random and inspected. Let X denote the number inspected that have improperly fitted doors.

(a) Find the density for X.

(b) Find $E[X]$ and $Var[X]$.

(c) Find the probability that at most one will be found with improperly fitted doors.

Example (Solution)

1. X follows a hypergeometric distribution with $N = 15$, $r = 4$ and $n = 3$. The density function is $f(x) = \frac{\binom{4}{x}\binom{11}{3-x}}{\binom{15}{3}}$ for $x = 0, 1, 2, 3$
2. $E[X] = \frac{12}{15} = 0.8$ and $Var[X] = \frac{3(4)(15-4)(15-3)}{15^2(14)} = 0.5029$.
3. $P(X \leq 1) = P(X = 0) + P(X = 1) = f(0) + f(1) = \frac{\binom{11}{3}}{455} + \frac{4\binom{11}{2}}{455} = 0.8462$

Poisson distribution

- The Poisson distribution is the discrete probability distribution of the number of events occurring in a given interval of time, given the average number of times the event occurs over that time period.
- The density function of a Poisson distribution is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

  for $x = 0, 1, 2, \ldots$ and $\lambda > 0$. λ is called the parameter of the distribution and is the average number of occured events in a unit of time. Notation: $X \in Poiss(\lambda)$
- $E[X] = Var[X] = \lambda$.
- If $X_1 \in Poiss(\lambda_1)$ and $X_2 \in Poiss(\lambda_2)$ then $X_1 + X_2 \in Poiss(\lambda_1 + \lambda_2)$.

Example

Let X be the number of typos on a printed page with a mean of 3 typos per page.

1. What is the probability that a randomly selected page has at least one typo on it?
2. What is the probability that three randomly selected pages have more than eight typos on it?

Solution:

1. $P(X \geq 1) = 1 - P(X = 0) = 1 - f(0) = 1 - e^{-3}$.

2. In this case $\lambda = 9$ since we have in average 9 typos on three printed pages.

$P(X > 8) = 1 - P(X \leq 8) = 1 - 0.456$ by table II page 692.

Example

Let X be the number of typos on a printed page with a mean of 3 typos per page.

1. What is the probability that a randomly selected page has at least one typo on it?
2. What is the probability that three randomly selected pages have more than eight typos on it?

Solution:

1. $P(X \geq 1) = 1 - P(X = 0) = 1 - f(0) = 1 - e^{-3}$.

2. In this case $\lambda = 9$ since we have in average 9 typos on three printed pages.

$P(X > 8) = 1 - P(X \leq 8) = 1 - 0.456$ by table II page 692.

Gamma distribution

- Gamma function: $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz$.
- A random variable X with density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

for $x > 0$, $\beta > 0$ and $\alpha > 0$, is said to have a gamma distribution with parameters α and β.

χ^2-squared distribution

- A continuous random variable is said to have a χ^2-distribution with γ degrees of freedom if X has a gamma distribution with parameters $\beta = 2$ anf $\alpha = \frac{\gamma}{2}$.
- $E[X] = \gamma$ and $Var[X] = 2\gamma$.
- Table IV p.695-696 gives the cumulative probabilities of χ^2-distributions. Note that the probabilities appear as the row heading (unlike other tables where the probabilities are the core of the tables.)
- If $X_1, \ldots, X_n$ have standard normal distributions and are independent, then $X_1^2 + \cdots + X_n^2$ follows a χ^2-distribution with n degrees of freedom.

Exponential distribution

- A continuous random variable follows an exponential distribution with parameter $\beta > 0$ if its density function is given by

$$f(x) = \frac{1}{\beta} e^{-x/\beta}$$

for $x > 0$, or equivalently $f(x) = \lambda e^{-\lambda x}$ where $\lambda = \frac{1}{\beta}$.

- $E[X] = \beta$ and $Var[X] = \beta^2$.
- The cumulative distribution function is given by

$$F(x) = 1 - e^{-\lambda x}.$$

Theorem

Let X be a random vairable with a Poisson distribution of parameter λ. Let W be the time of the occurence of the fiirst event. Then W has an exponetial distribution with parameter $\beta = \frac{1}{\lambda}$.

Example

Students arrive at a local bar and restaurant according to an approximate Poisson process at a mean rate of 30 students per hour. What is the probability that the bouncer has to wait more than 3 minutes to card the next student?

Solution The mean is 30 students per hour which is equivalent to $\frac{1}{2}$ student per minute. Therefore $\lambda = \frac{1}{2}$.

Let W be the time the bouncer has to wait. W follows an exponential distribution with parameter $\beta = 2$. Hence,

$$P(X > 3) = \int_3^\infty \frac{1}{2} e^{-x/2} = \left[-e^{-x/2}\right]_3^\infty = e^{-1.5}$$

Discrete distributions:

1. Geometric distribution
2. Binomial distribution
3. Hypergeometric distribution
4. Negative binomial distribution
5. Poisson distribution

Continuous distribution

1. Normal distribution
2. Exponential distribution
3. χ^2-distribution