Distributions

Review of discrete probability

Random variable

A random variable assigns a probability to a numerical outcome.

Review of discrete probability

Random variable

A random variable assigns a probability to a numerical outcome.

Expected value $E(x)$

$E(x) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + ... + x_n P(x_n) = \sum_{i=1}^{n} x_i P(x_i)$.

Review of discrete probability

Random variable

A random variable assigns a probability to a numerical outcome.

Expected value $E(x)$

$E(x) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + ... + x_n P(x_n) = \sum_{i=1}^{n} x_i P(x_i)$.

Variance $V(x)$

$$
\begin{aligned}
V(x) &= (x_1 - \mu)^2 \cdot P(x_1) + (x_2 - \mu)^2 \cdot P(x_2) + ... + (x_n - \mu)^2 P(x_n) \\
&= \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i) \\
\sigma_x &= \sqrt{V(x)}
\end{aligned}
$$

Continuous random variables

Random continuous variable

A random continuous variable x assigns a probability to a range of numerical outcomes $[min, max]$ on the real number line.

Continuous random variables

Random continuous variable

A random continuous variable x assigns a probability to a range of numerical outcomes $[min, max]$ on the real number line.

Probability density function (PDF)

A random continuous variable is described by the nonnegative continuous function f such that

$$\int_{min}^{max} f(x)\,\mathrm{d}x = 1 \quad \text{and} \quad P(a \leq x \leq b) = \int_{a}^{b} f(x)\,\mathrm{d}x$$

Continuous random variables

Random continuous variable

A random continuous variable x assigns a probability to a range of numerical outcomes $[min, max]$ on the real number line.

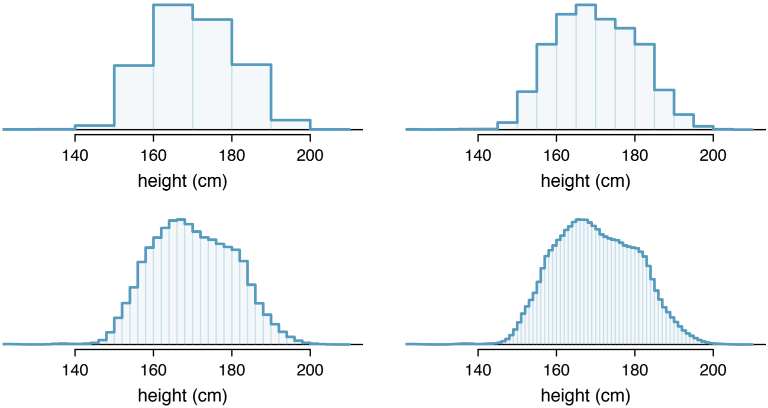
Probability density function (PDF)

A random continuous variable is described by the nonnegative continuous function f such that

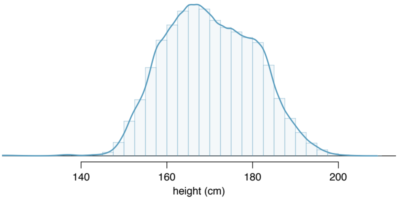
$$\int_{min}^{max} f(x)\,\mathrm{d}x = 1 \quad \text{and} \quad P(a \leq x \leq b) = \int_{a}^{b} f(x)\,\mathrm{d}x$$

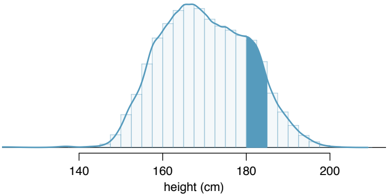
Example: $f(x) = x^3 - 2x^2 + x$

Integrate: $6[\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}] = 6(\frac{1}{4} - \frac{2}{3} + \frac{1}{2}) = 1$



Source: Diez *et al.* 2011





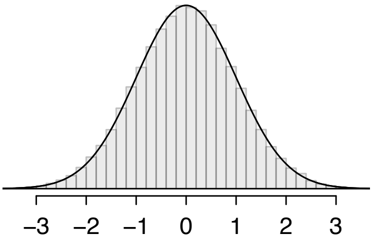
$$\int_{min}^{max} f(x)\,\mathrm{d}x = 1$$

$$P(180 \leq x \leq 185)$$

Source: Diez *et al.* 2011

Normal distribution

The normal distribution is a symmetric, unimodal distribution.



Notation $\mathcal{N}(\mu, \sigma)$

The standard normal distribution of mean 0 and standard deviation 1 is denoted $\mathcal{N}(0, 1)$, where μ and σ are the distribution parameters.

Source: Diez *et al.* 2011

Normal distribution

Standardised scores

The standardised score, or "Z-score", of an observation x, is equal to its distance from the mean divided by the standard deviation of the distribution: $Z = \frac{x-\mu}{\sigma}$.

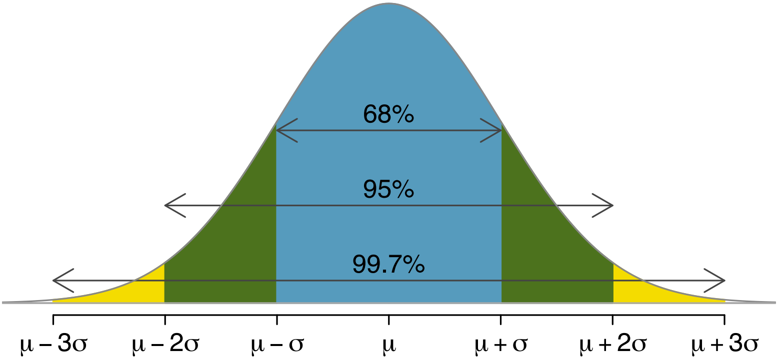
Use of Z-scores

- Determine how far an observation is from the mean
- Determine how far an observation is from the distribution

Example

Body Mass Index in the United States, 2009 (Stata).

Probability of falling within $|z| = \{1, 2, 3\}$



Source: Diez *et al.* 2011

Bernoulli distribution

Binary probabilistic events like *successive* coin flips ($S = \{0, 1\}$) are called “Bernoulli trials” and follow a geometric distribution.

Notation of proportion $\hat{p}$

For a sample series of Bernouilli trials, the estimated proportion of success is denoted

$$\hat{p} = \frac{n \text{ successes}}{N \text{ trials}} = \frac{0 + 0 + 1 + 0 + 1 +, ..., +0 + 1}{N}$$

The estimated proportion of failure is, conversely, $q = 1 - p$.

Demonstration

Show that a Bernoulli variable has a mean (or expected value) of $\mu = p$ and a standard deviation of $\sigma = \sqrt{p(1 - p)}$.

Geometric distribution

If p and $1 - p$ are the probabilities of success and failure in a Bernoulli trial, the underlying distribution has the following mean, variance and standard deviation:

Geometric distribution

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{1-p}{p^2} \quad \sigma = \sqrt{\frac{1-p}{p}}$$

The probability of success after n trials is $(1-p)^{n-1}p$ and decreases exponentially.

Exercise

Show the above by calculating the probability of finding a flatmate if everyone you ask has a 30% chance to accept. How likely is it that you will end with the first, second or third person you ask?

Solutions

Bernoulli variable

$$\begin{aligned} \mu &= E(x) = 0 \cdot P(x=0) + 1 \cdot P(x=1) \\ &= 0(1-p) + p = p \\ \sigma^2 &= P(x=0)(0-p)^2 + P(x=1)(1-p)^2 \\ &= (1-p)p^2 + p(1-p)^2 = p(1-p) \end{aligned}$$

Solutions

Bernoulli variable

$$\begin{aligned}\mu &= E(x) = 0 \cdot P(x=0) + 1 \cdot P(x=1) \\ &= 0(1-p) + p = p \\ \sigma^2 &= P(x=0)(0-p)^2 + P(x=1)(1-p)^2 \\ &= (1-p)p^2 + p(1-p)^2 = p(1-p)\end{aligned}$$

Flatmate search

The first person accepts (success) with a probability of $p = .3$, so the first trial fails $1 - p = 1 - .3 = 70\%$ of the time.
The probability that the second person accepts is $(1 - .3)(.3)$, the probability of the third $(1 - .3)(1 - .3)(.3)$ and so on.

Binomial distribution

When the Bernoulli trials are conducted simultaneously, the random variable follows a binomial distribution.

Relationship to combinations

The probability of getting k successes out of n trials is equivalent to drawing k elements from a sample space $\{0, 1\}$ of n elements: $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Binomial distribution

When the Bernoulli trials are conducted simultaneously, the random variable follows a binomial distribution.

Relationship to combinations

The probability of getting k successes out of n trials is equivalent to drawing k elements from a sample space $\{0, 1\}$ of n elements: $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Probability of a single success k out of n trials

$p^k(1-p)^{n-k}$

Binomial distribution

When the Bernoulli trials are conducted simultaneously, the random variable follows a binomial distribution.

Relationship to combinations

The probability of getting k successes out of n trials is equivalent to drawing k elements from a sample space $\{0, 1\}$ of n elements: $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Probability of a single success k out of n trials

$p^k(1-p)^{n-k}$

Binomial distribution

Probability of observing k successes in n independent trials:
$P(\text{trials}) \cdot P(\text{single success}) = \binom{n}{r} p^k (1-p)^{n-k}$

Binomial distribution

Mean, variance and standard deviation

$\mu = np \quad \sigma^2 = np(1-p) \quad \sigma = \sqrt{np(1-p)}$

Binomial distribution

Mean, variance and standard deviation

$\mu = np \quad \sigma^2 = np(1-p) \quad \sigma = \sqrt{np(1-p)}$

If 20% of people accept to donate \$10 to charities when asked, what is the probability of asking $n = 4$ people and getting exactly \$10 (single success)?

Binomial distribution

Mean, variance and standard deviation

$\mu = np \quad \sigma^2 = np(1-p) \quad \sigma = \sqrt{np(1-p)}$

If 20% of people accept to donate \$10 to charities when asked, what is the probability of asking $n = 4$ people and getting exactly \$10 (single success)?

$P(A = no) \cdot P(B = no) \cdot P(C = no) \cdot P(D = donate) =$
$(.8)(.8)(.8)(.2) = (.8)^3(.2) = (1-p)^{n-k}p^k$

Binomial distribution

Mean, variance and standard deviation

$\mu = np \quad \sigma^2 = np(1-p) \quad \sigma = \sqrt{np(1-p)}$

If 20% of people accept to donate \$10 to charities when asked, what is the probability of asking $n = 4$ people and getting exactly \$10 (single success)?

$P(A = no) \cdot P(B = no) \cdot P(C = no) \cdot P(D = donate) =$
$(.8)(.8)(.8)(.2) = (.8)^3(.2) = (1-p)^{n-k}p^k$

If 30% of smokers die from lung cancer, what is the likelihood of none dying that way in a group of 3 random smokers?

Binomial distribution

Mean, variance and standard deviation

$\mu = np \quad \sigma^2 = np(1-p) \quad \sigma = \sqrt{np(1-p)}$

If 20% of people accept to donate \$10 to charities when asked, what is the probability of asking $n = 4$ people and getting exactly \$10 (single success)?

$P(A = no) \cdot P(B = no) \cdot P(C = no) \cdot P(D = donate) =$
$(.8)(.8)(.8)(.2) = (.8)^3(.2) = (1-p)^{n-k}p^k$

If 30% of smokers die from lung cancer, what is the likelihood of none dying that way in a group of 3 random smokers?

Mean probability: $\mu = .3 \times 4 = 1.2$ will die on average.
$P(k = 0) = \binom{3}{0}(.3)^0(1-.3)^{3-0} = \frac{3!}{0!\,(3-0)!}(.3)^0(.7)^3$
There is a $1 \times 1 \times (.7)^3 \approx 34\%$ likelihood of none dying that way.

Homework

Read Evans *et al.* 2012, ch. 4–5* for next week
and enjoy the rest of your day.

For next week, make sure that you know how to work out the exercises shown in slides 7–8.

* http://www.ck12.org/book/CK-12-Advanced-Probability-and-Statistics-Concepts/