Probability: Review exercises

Things to know

Definitions

- Probability, sample space and events
- Discrete and continuous random variables
- Mean, variance and standard deviations

Distributions

- Normal distribution and standardized scores
- Bernoulli trials and geometric distribution
- Binomial distribution

Discrete probability

Probability of a random variable x

$P(x) = \frac{n(x)}{n(S)} \quad 0 \leq P \leq 1$

Expected value $E(x)$, or mean μ

The mean measures the average numerical outcome of x.
$E(x) = \sum_{i=1}^{n} x_i P(x_i)$.

Variance $V(x)$ or σ^2, and standard deviation σ

Variability measures the squared sum of deviations from the mean.
$V(x) = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i) \quad \sigma_x = \sqrt{V(x)}$

Continuous variables

Probability density function of x

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

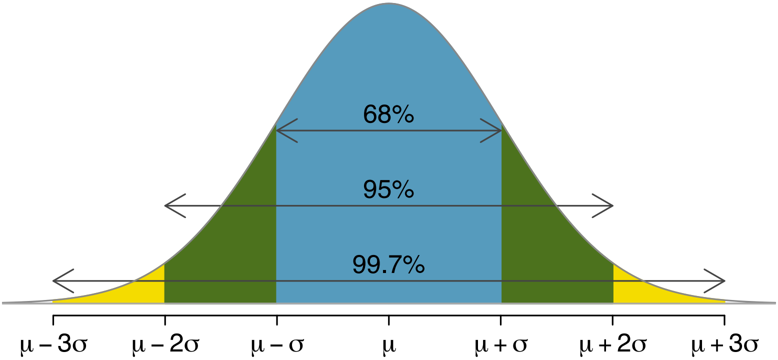
Standard normal distribution $\mathcal{N}(0,1)$

- approx. 68% of values at $\mu \pm 1\sigma$
- approx. 95% of values at $\mu \pm 2\sigma$
- approx. 99% of values at $\mu \pm 3\sigma$

Standardized score

$Z = \frac{x-\mu}{\sigma}$

Standard normal distribution



Source: Diez *et al.* 2011

Probability distributions

1. Identify the distribution

- The binomial random variable X is the number of successes over n independent trials.
- The geometric distribution for Bernoulli trials focuses on how many trials produce a single success.
- The Poisson and negative binomial distributions have other focused uses and are not covered.

2. Describe the distribution

- Start by mentioning its name: "the binomial probability of..."
- Write up the formula for its random variable X.
- Compute simple probabilities, approximate otherwise

Bernoulli trials: successes over a dichotomous outcome

Bernoulli variable

p is the proportion of successes in $S = \{0, 1\}$
$\mu = p \quad \sigma = \sqrt{p(1-p)}$

Probability of a single success after n trials

$(1-p)^{n-1}p$ (exponentially decreasing geometric distribution)

Mean, variance and standard deviation

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

Binomial distribution: n independent Bernoulli trials

Probability of a single success k out of n trials

$$P(x = 1) = p^k (1 - p)^{n-k}$$

Probability of k successes

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Mean, variance and standard deviation

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

Normal approximation

If np and $n(1 - p)$ are both at least 10, the approximate normal distribution has parameters corresponding to the mean and standard deviation of the binomial distribution.

Exercise with discrete probabilities

Exercise 1: Trader gains

A trader wins \$300 on 90% of his transactions and loses \$1,000 on 10% of them. (a) How much does he gain on average? (b) At what probability is his expected gain null?

Exercise with discrete probabilities

Exercise 1: Trader gains

A trader wins \$300 on 90% of his transactions and loses \$1,000 on 10% of them. (a) How much does he gain on average? (b) At what probability is his expected gain null?

$$
\begin{aligned}
E(x) &= \sum xp(x) \\
&= (.9)(300) + (.1)(-1000) \\
&= 270 - 100 = 170
\end{aligned}
$$

(a) The trader expects to gain \$170 per transaction on average.
(b) $E(x) = 0$ when $300(p) = 1000(1-p)$, i.e. when $p = \frac{10}{13} \approx .77$.

Exercises with discrete probabilities

Exercise 2: Sexual transmission

If 40% of a population is contaminated with a sexual disease and sex acts occur at random, what percentage of sex acts poses a contamination risk?

Exercises with discrete probabilities

Exercise 2: Sexual transmission

If 40% of a population is contaminated with a sexual disease and sex acts occur at random, what percentage of sex acts poses a contamination risk?

$$P(x_1 = 0, x_2 = 0) = (.6)(.6) = .36$$
$$P(x_1 = 1, x_2 = 1) = (.4)(.4) = .16$$
$$P(x_1 = 0, x_2 = 1) = (.6)(.4) = .24$$
$$P(x_1 = 1, x_2 = 0) = (.4)(.6) = .24$$

If sex acts within the population are random, 48% of sex acts expose one of the partners (x_1, x_2) to contamination.

Exercise with continuous variables

Exercise 3: SAT scores (Diez *et al.* 2012)

SAT scores closely follow the normal model with mean $\mu = 1500$ and standard deviation $\sigma = 300$. (a) About what percent of test takers score 900 to 2100? (b) What percent score between 1500 and 2100?

Exercise with continuous variables

Exercise 3: SAT scores (Diez *et al.* 2012)

SAT scores closely follow the normal model with mean $\mu = 1500$ and standard deviation $\sigma = 300$. (a) About what percent of test takers score 900 to 2100? (b) What percent score between 1500 and 2100?

(a) 900 and 2100 represent two standard deviations above and below the mean, which means about 95% of test takers will score between 900 and 2100.

(b) Since the normal model is symmetric, then half of the test takers from part (a) ($95\% = 47.5\%$ of all test takers) will score 900 to 1500 while 47.5% score between 1500 and 2100.

Exercises with Bernoulli trials

Exercise 4: Imitation cascade

If there is exactly one chance out of 10,000 that a revolution occuring at time t_0 will be imitated in a neighbour country at time t, at what time will it most likely be imitated?

Exercises with Bernoulli trials

Exercise 4: Imitation cascade

If there is exactly one chance out of 10,000 that a revolution occuring at time t_0 will be imitated in a neighbour country at time t, at what time will it most likely be imitated?

The probability of imitation at $t = 1$ is .0001. At $t = 2$, that probability becomes (.9999)(.0001) and decreases exponentially with t as $P(t) = (1 - p)^{t-1}p$ in a geometric distribution.

The most likely time for imitation is immediately after the event when the probability of success p is constant. In reality, however, p varies and the event is not a Bernoulli trial.

Exercises with Bernoulli trials

Exercise 5: Equation review

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

Exercises with Bernoulli trials

Exercise 5: Equation review

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

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

Bernoulli trials are 'one dimension' of binomial distributions, which are equivalent to running several independent Bernoulli trials; for a binomial distribution, $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

Exercises with binomial probabilities

Exercise 6: Collecting donations

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?

Exercises with binomial probabilities

Exercise 6: Collecting donations

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?

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

Exercises with binomial probabilities

Exercise 7: Getting married (from Evans et al., p. 148)

If 6% of unmarried women of 35 years of age will marry within five years, calculate the probabilities that out of a random sample of 9 unmarried women of that age, (a) none will get married within 5 years. (b) at least one will get married within 5 years.

Exercises with binomial probabilities

Exercise 7: Getting married (from Evans et al., p. 148)

If 6% of unmarried women of 35 years of age will marry within five years, calculate the probabilities that out of a random sample of 9 unmarried women of that age, (a) none will get married within 5 years. (b) at least one will get married within 5 years.

$$P(k) = \binom{9}{k}(.06)^k(1-.06)^{9-k}$$
$$P(0) = 1 \cdot 1 \cdot (.94)^9 \approx .57$$
$$P(x > 0) = 1 - P(0) \approx .42$$

Exercises with binomial probabilities

Exercise 8: Probability sampling

Assume a sample of 50 people randomly selected from a population where 20% oppose gay marriage. (a) Calculate the mean and standard deviation of opposition to gay marriage in the sample. (b) How likely is it that the sample contains a maximum of two opponents to gay marriage?

Exercises with binomial probabilities

Exercise 8: Probability sampling

Assume a sample of 50 people randomly selected from a population where 20% oppose gay marriage. (a) Calculate the mean and standard deviation of opposition to gay marriage in the sample. (b) How likely is it that the sample contains a maximum of two opponents to gay marriage?

$$\mu = np = (50)(.2) = 10$$
$$\sigma^2 = np(1-p) = (50)(.2)(.8) = 8 \quad \sigma = \sqrt{8} \approx 3$$

(a) The sample will contain 10 opponents on average.

(b) It seems highly unlikely, given $\mu = 10$ and $\sigma = 3$, but the normal approximation is unreliable here because $np(1-p) < 10$.

Exercises with binomial probabilities

Exercise 9: Survey response rates

Assume a survey with a 10% response rate. If 15,000 households are contacted, is it likely to expect that 1,530 will agree to respond?

Exercises with binomial probabilities

Exercise 9: Survey response rates

Assume a survey with a 10% response rate. If 15,000 households are contacted, is it likely to expect that 1,530 will agree to respond?

$$\mu = np = (15000)(.1) = 1500$$

$$\sigma^2 = np(1-p) = (15000)(.1)(.9) = 1350 \quad \sigma = \sqrt{1350} \approx 36$$

The distribution approaches $\mathcal{N}(\mu = 1500, \sigma = 36)$, which makes a $P(x = 1530)$ a rather likely outcome.

Homework

Read CK-12 handbook ch. 7 for next week,

prepare for a short exam (pen-and-paper, 45 min.),
check your ENTG to know at what time the exam is for you

and enjoy the rest of your day.