

Review problems with solutions
Class 21, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Summary

- Data: x_1, \dots, x_n
 - Basic statistics: sample mean, sample variance, sample median
 - Likelihood, maximum likelihood estimate (MLE)
 - Bayesian updating: prior, likelihood, posterior, predictive probability, probability intervals; prior and likelihood can be discrete or continuous
 - NHST: H_0 , H_A , significance level, rejection region, power, type 1 and type 2 errors, p -values.
-

2 Basic statistics

Data: x_1, \dots, x_n .

$$\text{sample mean} = \bar{x} = \frac{x_1 + \dots + x_n}{n}$$

$$\text{sample variance} = s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

sample median = middle value

Example. Data: 1, 2, 3, 6, 8.

$$\bar{x} = 4, \quad s^2 = \frac{9+4+1+4+16}{4} = 8.5, \quad \text{median} = 3.$$

3 Likelihood

x = data

θ = parameter of interest or hypotheses of interest

Likelihood:

$p(x | \theta)$ (discrete distribution)

$f(x | \theta)$ (continuous distribution)

Log likelihood :

$$\ln(p(x | \theta)).$$

$$\ln(f(x | \theta)).$$

Likelihood examples. Find the likelihood function of each of the following.

1. Coin with probability of heads θ . Toss 10 times get 3 heads.
2. Wait time follows $\exp(\lambda)$. In 5 trials wait 3,5,4,5,2
3. Usual 5 dice. Two rolls, 9, 5. (Likelihood given in a table)
4. $x_1, \dots, x_n \sim N(\mu, \sigma^2)$
5. $x = 6$ drawn from uniform(0, θ)
6. $x \sim \text{uniform}(0, \theta)$

Solutions.

1. Let x be the number of heads in 10 tosses. $P(x = 3 | \theta) = \binom{10}{3} \theta^3 (1 - \theta)^7$.
2. $f(\text{data} | \lambda) = \lambda^5 e^{-\lambda(3+5+4+5+2)} = \lambda^5 e^{-19\lambda}$

3.

Hypothesis θ	Likelihood $P(\text{data} \theta)$
4-sided	0
6-sided	0
8-sided	0
12-sided	1/144
20-sided	1/1400

$$4. P(\text{data} | \mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{[(x_1-\mu)^2 + (x_2-\mu)^2 + \dots + (x_n-\mu)^2]}{2\sigma^2}}$$

$$5. f(x = 6 | \theta) = \begin{cases} 0 & \text{if } \theta < 6 \\ 1/\theta & \text{if } 6 \leq \theta \end{cases}$$

$$6. f(x | \theta) = \begin{cases} 0 & \text{if } \theta < x \text{ or } x < 0 \\ 1/\theta & \text{if } 0 \leq x \leq \theta \end{cases}$$

3.1 Maximum likelihood estimates (MLE)

Methods for finding the maximum likelihood estimate (MLE).

- Discrete hypotheses: compute each likelihood
- Discrete hypotheses: maximum is obvious
- Continuous parameter: compute derivative (often use log likelihood)
- Continuous parameter: maximum is obvious

Examples. Find the MLE for each of the examples in the previous section.

Solutions.

1. $\ln(f(x-3 | \theta)) = \ln\binom{10}{3} + 3\ln(\theta) - 7\ln(1-\theta)$.

Take the derivative and set to 0: $\frac{3}{\theta} + \frac{7}{1-\theta} = 0 \Rightarrow \hat{\theta} = \frac{3}{10}$.

2. $\ln(f(\text{data} | \lambda)) = 5\ln(\lambda) - 19\lambda$.

Take the derivative and set to 0: $\frac{5}{\lambda} - 19 = 0 \Rightarrow \hat{\lambda} = \frac{5}{19}$.

3. Read directly from the table: MLE = 12-sided die.

4. For the exam do not focus on the calculation here. You should understand the idea that we need to set the partial derivatives with respect to μ and σ to 0 and solve for the critical point $(\hat{\mu}, \hat{\sigma}^2)$.

The result is $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{\sum(x_i - \hat{\mu})^2}{n}$.

5. Because of the term $1/\theta$ in the likelihood. The likelihood is at a maximum when θ is as small as possible. **answer:** $\hat{\theta} = 6$.

6. This is identical to problem 5 except the exact value of x is not given. **answer:** $\hat{\theta} = x$.

4 Bayesian updating

4.1 Bayesian updating: discrete prior-discrete likelihood.

Jon has 1 four-sided, 2 six-sided, 2 eight-sided, 2 twelve sided, and 1 twenty-sided dice. He picks one at random and rolls a 7.

1. For each type of die, find the posterior probability Jon chose that type.
2. What are the posterior odds Jon chose the 20-sided die?
3. Compute the prior predictive probability of rolling a 7 on the first roll.
4. Compute the posterior predictive probability of rolling an 8 on the second roll.

Solutions.

1.. Make a table. (We include columns to answer question 4.)

Hypothesis θ	Prior $P(\theta)$	Likelihood $f(x_1 = 7 \theta)$	Unnorm. posterior	posterior $f(\theta x_1 = 7)$	likelihood $P(x_2 = 8 \theta)$	unnorm. posterior
4-sided	1/8	0	0	0	0	0
6-sided	1/4	0	0	0	0	0
8-sided	1/4	1/8	1/32	1/32c	1/8	1/256c
12-sided	1/4	1/12	1/48	1/48c	1/12	1/576c
20-sided	1/8	1/20	1/160	1/160c	1/20	1/3200c
Total	1		$c = \frac{1}{32} + \frac{1}{48} + \frac{1}{160}$	1		

The posterior probabilities are given in the 5th column of the table. The total probability $c = \frac{7}{120}$ is also the answer to problem 3.

2. Odds(20-sided | $x_1 = 7$) = $\frac{P(20\text{-sided} | x_1=7)}{P(\text{not } 20\text{-sided} | x_1=7)} = \frac{1/160c}{1/32c+1/48c} = \frac{1/160}{5/96} = \frac{96}{800} = \frac{3}{25}$.
 3. $P(x_1 = 7) = c = 7/120$.
 4. See the last two columns in the table. $P(x_2 = 8 | x_1 = 7) = \frac{1}{256c} + \frac{1}{576c} + \frac{1}{3200c} = \frac{49}{480}$.
-

4.2 Bayesian updating: conjugate priors.

Beta prior, binomial likelihood

Data: $x \sim \text{binomial}(n, \theta)$. θ is unknown.

Prior: $f(\theta) \sim \text{beta}(a, b)$

Posterior: $f(\theta | x) \sim \text{beta}(a + x, b + n - x)$

1. Suppose $x \sim \text{binomial}(30, \theta)$, $x = 12$. If we have a prior $f(\theta) \sim \text{beta}(1, 1)$ find the posterior for θ .

Beta prior, geometric likelihood

Data: x

Prior: $f(\theta) \sim \text{beta}(a, b)$

Posterior: $f(\theta | x) \sim \text{beta}(a + x, b + 1)$.

2. Suppose $x \sim \text{geometric}(\theta)$, $x = 6$. If we have a prior $f(\theta) \sim \text{beta}(4, 2)$ find the posterior for θ .

Normal prior, normal likelihood

$$a = \frac{1}{\sigma_{\text{prior}}^2} \qquad b = \frac{n}{\sigma^2}$$

$$\mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a + b}, \qquad \sigma_{\text{post}}^2 = \frac{1}{a + b}.$$

3. In the population IQ is normally distributed: $\theta \sim N(100, 15^2)$. An IQ test finds a person's 'true' IQ + random error $\sim N(0, 10^2)$. Someone takes the test and scores 120.

Find the posterior pdf for this person's IQ.

Solutions.

1. $f(\theta) \sim \text{beta}(1, 1)$, $x \sim \text{binom}(30, \theta)$. $x = 12$, so $f(\theta | x = 12) \sim \beta(13, 19)$
2. $f(\theta) \sim \text{beta}(4, 2)$, $x \sim \text{geom}\theta$. $x = 6$, so $f(\theta | x = 6) \sim \beta(10, 3)$
3. Prior, $f(\theta) \sim N(100, 15^2)$, $x \sim N(\theta, 10^2)$.

So we have, $\mu_{\text{prior}} = 100$, $\sigma_{\text{prior}}^2 = 15^2$, $\sigma^2 = 10^2$, $n = 1$, $\bar{x} = x = 120$.

Applying the normal-normal update formulas: $a = \frac{1}{15^2}$, $b = \frac{1}{10^2}$. This gives

$$\mu_{\text{post}} = \frac{100/15^2 + 120/10^2}{1/15^2 + 1/10^2} = 113.8, \quad \sigma_{\text{post}}^2 = \frac{1}{1/15^2 + 1/10^2} = 69.2$$

Bayesian updating: continuous prior-continuous likelihood

Examples. Update from prior to posterior for each of the following with the given data. Graph the prior and posterior in each case.

- Romeo is late: likelihood: $x \sim U(0, \theta)$, prior: $U(0, 1)$, data: 0.3, 0.4, 0.4.
- Waiting times: likelihood: $x \sim \exp(\lambda)$, prior: $\lambda \sim \exp(2)$, data: 1, 2.
- Waiting times: likelihood: $x \sim \exp(\lambda)$, prior: $\lambda \sim \exp(2)$, data: x_1, x_2, \dots, x_n .

Solutions.

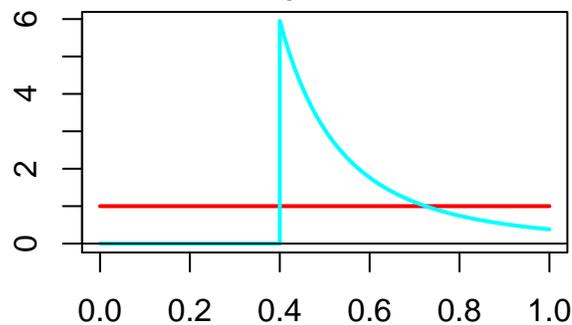
- In the update table we split the hypotheses into the two different cases $\theta < 0.4$ and

	prior	likelihood	unnormalized	posterior
hyp.	$f(\theta)$	$f(\text{data} \theta)$	posterior	$f(\theta \text{data})$
$\theta < 0.4$	$d\theta$	0	0	0
$\theta \geq 0.4$	$d\theta$	$\frac{1}{\theta^3}$	$\frac{d\theta}{\theta^3}$	$\frac{1}{T\theta^3} d\theta$
Tot.	1		T	1

The total probability

$$T = \int_{0.4}^1 \frac{d\theta}{\theta^3} \Rightarrow T = -\frac{1}{2\theta^2} \Big|_{0.4}^1 = \frac{21}{8} = 2.625.$$

We use $1/T$ as a normalizing factor to make the total posterior probability equal to 1.

Prior and posterior for θ 

Prior in red, posterior in cyan

- This follows the same pattern as problem 1.

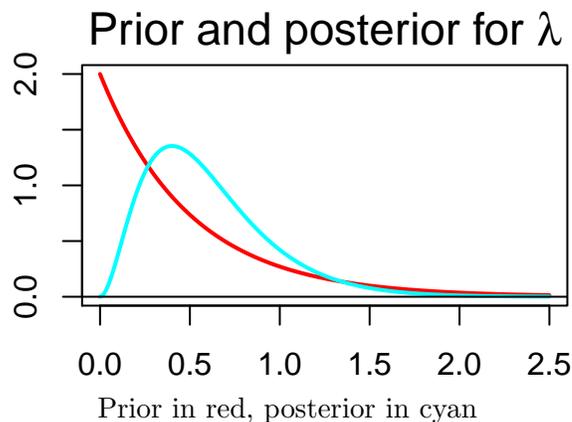
The likelihood $f(\text{data} | \lambda) = \lambda e^{-\lambda \cdot 1} \lambda e^{-\lambda \cdot 2} = \lambda^2 e^{-3\lambda}$.

	prior	likelihood	unnormalized	posterior
hyp.	$f(\lambda)$	$f(\text{data} \lambda)$	posterior	$f(\lambda \text{data})$
$0 < \lambda < \infty$	$2e^{-2\lambda}$	$\lambda^2 e^{-3\lambda}$	$2\lambda^2 e^{-5\lambda} d\lambda$	$\frac{2}{T} \lambda^2 e^{-5\lambda} d\lambda$
Tot.	1		T	1

The total probability (computed using integration by parts)

$$T = \int_0^{\infty} 2\lambda^2 e^{-5\lambda} d\lambda \Rightarrow T = \frac{4}{125}.$$

We use $1/T$ as a normalizing factor to make the total posterior probability equal to 1.



3. This is nearly identical to problem 2 except the exact values of the data are not given, so we have to work abstractly.

The likelihood $f(\text{data} \mid \lambda) = \lambda^n e^{-\lambda \sum x_i}$.

hyp.	prior $f(\lambda)$	likelihood $f(\text{data} \mid \lambda)$	unnormalized posterior	posterior $f(\lambda \mid \text{data})$
$0 < \lambda < \infty$	$2e^{-2\lambda}$	$\lambda^n e^{-\lambda \sum x_i}$	$2\lambda^n e^{-\lambda(2+\sum x_i)} d\lambda$	$\frac{2}{T} \lambda^n e^{-\lambda(2+\sum x_i)} d\lambda$
Tot.	1		T	1

For this problem you should be able to write down the integral for the total probability. We won't ask you to compute something this complicated on the exam.

$$T = \int_0^{\infty} 2\lambda^n e^{-\lambda \sum x_i} d\lambda \Rightarrow T = \frac{2 * n!}{(2 + \sum x_i)^{n+1}}.$$

We use $1/T$ as a normalizing factor to make the total posterior probability equal to 1.

The plot for problem 2 is one example of what the graphs can look like.

5 Null hypothesis significance testing (NHST)

5.1 NHST: Steps

1. Specify H_0 and H_A .
2. Choose a significance level α .
3. Choose a test statistic and determine the null distribution.
4. Determine how to compute a p -value and/or the rejection region.
5. Collect data.
6. Compute p -value or check if test statistic is in the rejection region.
7. Reject or fail to reject H_0 .

Make sure you can use the probability tables.

5.2 NHST: One-sample t -test

- Data: we assume normal data with both μ and σ unknown:

$$x_1, x_2, \dots, x_n \sim N(\mu, \sigma^2).$$

- Null hypothesis: $\mu = \mu_0$ for some specific value μ_0 .

- Test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- Null distribution: $t(n-1)$

Example. z and one-sample t -test

For both problems use significance level $\alpha = .05$.

Assume the data 2, 4, 4, 10 is drawn from a $N(\mu, \sigma^2)$.

Take $H_0: \mu = 0$; $H_A: \mu \neq 0$.

1. Assume $\sigma^2 = 16$ is known and test H_0 against H_A .
2. Now assume σ^2 is unknown and test H_0 against H_A .

Solutions.

We have $\bar{x} = 5$, $s^2 = \frac{9+1+1+25}{3} = 12$

1. We'll use \bar{x} for the test statistic (we could also use z). The null distribution for \bar{x} is $N(0, 4^2/4)$. This is a two-sided test so the rejection region is

$$(\bar{x} \leq \sigma_{\bar{x}} z_{.975} \text{ or } \bar{x} \geq \sigma_{\bar{x}} z_{.025}) = (-\infty, -3.9199] \cup [3.9199, \infty)$$

Since our sample mean $\bar{x} = 5$ is in the rejection region we reject H_0 in favor of H_A . Repeating the test using a p -value:

$$p = P(|\bar{x}| \geq 5 \mid H_0) = P\left(\frac{|\bar{x}|}{2} \geq \frac{5}{2} \mid H_0\right) = P(z \geq 2.5) = .012$$

Since $p < \alpha$ we reject H_0 in favor of H_A .

2. We'll use $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ for the test statistic. The null distribution for t is t_3 . For the data we have $t = 5/\sqrt{3}$. This is a two-sided test so the p -value is

$$p = P(|t| \geq 5/\sqrt{3} \mid H_0) = .06318$$

Since $p > \alpha$ we do not reject H_0 .

5.3 Two-sample t -test: equal variances

Data: we assume normal data with μ_x, μ_y and (same) σ unknown:

$$x_1, \dots, x_n \sim N(\mu_x, \sigma^2), \quad y_1, \dots, y_m \sim N(\mu_y, \sigma^2)$$

Null hypothesis H_0 : $\mu_x = \mu_y$.

$$\text{Pooled variance: } s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m} \right).$$

$$\text{Test statistic: } t = \frac{\bar{x} - \bar{y}}{s_p}$$

Null distribution: $f(t | H_0)$ is the pdf of $T \sim t(n+m-2)$

Example.

We have data from 1408 women admitted to a maternity hospital for (i) medical reasons or through (ii) unbooked emergency admission. The duration of pregnancy is measured in complete weeks from the beginning of the last menstrual period.

(i) Medical: 775 obs. with $\bar{x} = 39.08$ and $s^2 = 7.77$.

(ii) Emergency: 633 obs. with $\bar{x} = 39.60$ and $s^2 = 4.95$

1. Set up and run a two-sample t -test to investigate whether the duration differs for the two groups.

2. What assumptions did you make?

Solutions.

1. The pooled variance for this data is

$$s_p^2 = \frac{774(7.77) + 632(4.95)}{1406} \left(\frac{1}{775} + \frac{1}{633} \right) = .0187$$

The t statistic for the null distribution is

$$\frac{\bar{x} - \bar{y}}{s_p} = -3.8064$$

Rather than compute the two-sided p -value using `2*tcdf(-3.8064, 1406)` we simply note that with 1406 degrees of freedom the t distribution is essentially standard normal and 3.8064 is almost 4 standard deviations. So

$$P(|t| \geq 3.8064) = P(|z| \geq 3.8064)$$

which is very small, much smaller than $\alpha = .05$ or $\alpha = .01$. Therefore we reject the null hypothesis in favor of the alternative that there is a difference in the mean durations.

2. We assumed the data was normal and that the two groups had equal variances. Given the big difference in the sample variances this assumption might not be warranted.

Note: there are significance tests to see if the data is normal and to see if the two groups have the same variance.

5.4 Chi-square test for goodness of fit

Three treatments for a disease are compared in a clinical trial, yielding the following data:

	Treatment 1	Treatment 2	Treatment 3
Cured	50	30	12
Not cured	100	80	18

Use a chi-square test to compare the cure rates for the three treatments

Solution. The null hypothesis is $H_0 =$ all three treatments have the same cure rate.

Under H_0 the MLE for the cure rate is: (total cured)/(total treated) = $92/290 = .317$.

Given H_0 we get the following table of observed and expected counts. We include the fixed values in the margins

	Treatment 1	Treatment 2	Treatment 3	
Cured	50, 47.6	30, 34.9	12, 9.5	92
Not cured	100, 102.4	80, 75.1	18, 20.5	198
	150	110	30	

Pearson's chi-square statistic: $X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.13$.

Likelihood ratio statistic: $G = 2 \sum O_i \ln(O_i/E_i) = 2.12$.

Because the margins are fixed we can put values in 2 of the cells freely and then all the others are determined: degrees of freedom = 2. Using R we compute the p -value using the χ^2 distribution with 2 degrees of freedom.

$$p = 1 - \text{pchisq}(2.12, 2) = .346$$

(We used the G statistic, but we would get essentially the same answer using X^2 .)

For the exam you would have to use the χ^2 table to estimate the p -value. In the $df = 2$ row of the table 2.12 is between the critical values for $p = 0.3$ and $p = 0.5$.

The problem did not specify a significance level, but a p -value of .35 does not support rejecting H_0 at any common level. We do not conclude that the treatments have differing efficacy.

5.5 F -test = one-way ANOVA

Like t -test but for n groups of data with m data points each.

$$y_{i,j} \sim N(\mu_i, \sigma^2), \quad y_{i,j} = j^{\text{th}} \text{ point in } i^{\text{th}} \text{ group}$$

Assumptions: data for each group is an independent normal sample with (possibly) different means but the same variance.

Null-hypothesis is that means are all equal: $\mu_1 = \dots = \mu_n$

Test statistic is $\frac{MS_B}{MS_W}$ where:

$MS_B = \text{between group variance} = \frac{m}{n-1} \sum (\bar{y}_i - \bar{y})^2$

$MS_W = \text{within group variance} = \text{sample mean of } s_1^2, \dots, s_n^2$

Idea: If μ_i are equal, this ratio should be near 1.

Null distribution is F-statistic with $n-1$ and $n(m-1)$ d.o.f.:

$$\frac{MS_B}{MS_W} \sim F_{n-1, n(m-1)}$$

Example. The table shows recovery time in days for three medical treatments.

1. Set up and run an F-test.
2. Based on the test, what might you conclude about the treatments?

T_1	T_2	T_3
6	8	13
8	12	9
4	9	11
5	11	8
3	6	7
4	8	12

For $\alpha = .05$, the critical value of $F_{2,15}$ is 3.68.

Solution.

1. It's not stated but we have to assume independence and normality.

$n = 3$ groups, $m = 6$ data points in each group.

F-stat: $f \sim F_{n-1, n(m-1)} = F_{2,15}$.

Group means: (Treatments 1-3): $\bar{y}_1 = 5$, $\bar{y}_2 = 9$, $\bar{y}_3 = 10$.

Grand mean: $\bar{y} = 8$.

Group variances: $s_1^2 = 16/5$, $s_2^2 = 24/5$, $s_3^2 = 28/5$.

$MS_B = \frac{6}{2}(14) = 42$, $MS_W = \frac{68}{15}$, $f = \frac{MS_B}{MS_W} = \frac{42}{68/15} = 9.264$.

2. Since $9.264 > 3.68$, at a significance level of 0.05 we reject the null hypothesis that all the means are equal.

5.6 NHST: some key points

1. The significance level α is not the probability of being wrong overall. It's the probability of being wrong if the null hypothesis is true.
2. Likewise, power is not a probability of being right. It's the probability of being right if a particular alternate hypothesis is true.

MIT OpenCourseWare
<http://ocw.mit.edu>

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.