MIDTERM EXAM PINK SOLUTION

BMED 2400
Instructor Brani Vidakovic
TAs: Maria Amoreth Gozo and Nader Aboujamous
Thursday, October 28, 2010.

Name ______________________________

Problem	Eye Color	Nylon Fiber	Aerobic Capacity	True/False	Total
Score	/20	/35	/25	/20	/100

3.17 Eye Color. The eye color of a child is determined by a pair of genes, one coming from each parent. If **b** and **B** denote blue- and brown-eyed genes, then a child can inherit the following pairs: **bb, bB, Bb** and **BB**. The **B** gene is dominant, that is, the child will have brown eyes when the pairs are **Bb, bB** or **BB**, and blue eyes only for the **bb** combination. A parent passes to a child either gene from his/her pair with equal probabilities. [1]

Megan's parents are both brown-eyed but Megan has blue eyes. Megan's brown-eyed sister is pregnant and her husband has blue eyes.

What is the probability that the baby will have blue eyes.

Solution:

Since Megan has blue eyes and both parents are brown-eyed, then the parents are both **Bb**. Without any information on Megan sister's phenotype, the distribution of her allele pairs would be

	BB	**Bb**	**bb**
probs	1/4	1/2	1/4

However, since we know that Megan's sister has brown eyes, then the conditional probabilities are calculated as

$$P(\{\mathbf{BB}\}|\{\mathbf{BB,\ Bb}\}) = \frac{P(\{\mathbf{BB}\} \cap \{\mathbf{BB,\ Bb}\})}{P(\{\mathbf{BB,\ Bb}\})} = \frac{P(\{\mathbf{BB}\})}{P(\{\mathbf{BB,\ Bb}\})} = \frac{1/4}{3/4} = 1/3.$$

Similarly, $P(\{\mathbf{Bb}\}|\{\mathbf{BB,\ Bb}\}) = 2/3$ and $P(\{\mathbf{bb}\}|\{\mathbf{BB,\ Bb}\}) = 0$.

Thus, after information about Megan sister's phenotype her genotype distribution is

	BB	**Bb**	**bb**
probs	1/3	2/3	0

Megan sister's husband allays passes **b** allele, and the child will be blue-eyed only if Megan's sister passes allele **b**. This happens with probability

$$P(\{\text{Megan's sister is } \mathbf{Bb}\}) \times P(\{\mathbf{b} \text{ is passed from } \mathbf{Bb}\ ) = 2/3 \times 1/2 = \boxed{1/3}$$

[1] This description is simplified and in fact there are several genes affecting the eye color and the amount of yellow and black pigments in the iris, leading to shades of colors including green and hazel.

5.27 Silver Coated Nylon Fiber. Silver-coated nylon fiber is used in hospitals for its anti-static electricity properties, as well as for antibacterial and antimycotic effects. In the production of silver-coated nylon fibers, the extrusion process is interrupted from time to time by blockages occurring in the extrusion dyes. The time in hours between blockages, T, has an exponential $\mathcal{E}(\lambda)$ distribution, where $\lambda = 1/10$ is the rate parameter.

Find the probabilities that

(a) a run continues for at least 10 hours,

(b) a run lasts less than 15 hours.

You can use MATLAB's `expcdf` function, alternatively, since the expression for exponential cdf is simple, you may do direct evaluation. Be careful about parametrization of exponentials in MATLAB.

Suppose now that the rate parameter λ is unknown, but there are three measurements of inter-blockage times, $T_1 = 3$, $T_2 = 13$, and $T_3 = 8$.

(c) estimate parameter λ using the moment-matching procedure. Write down the likelihood.

(d) What is the Bayes estimator of λ if the prior is $\pi(\lambda) = \frac{1}{\sqrt{\lambda}}, \ \lambda > 0$.

Hint. In (d) the prior is not proper distribution, but the posterior is. Identify the posterior from the product of the likelihood from (c) and the prior.

Solution:

```
%Silver Coated Nylon Fibers

% (a)
1 - expcdf(10, 10) % 0.3679
% (b)
expcdf(15,10)  %0.7769
```

(c) If T is exponential $\mathcal{E}(\lambda)$ where λ is the rate parameter, then $ET = 1/\lambda$. The moment matching estimator is $\hat{\lambda}_{mm} = 1/\bar{T}$.

Here $\bar{T} = \frac{3+13+8}{3} = 8$, so $\hat{\lambda}_{mm} = 1/8 = 0.125$.

The likelihood is:

$$\lambda e^{-3\lambda} \times \lambda e^{-13\lambda} \times \lambda e^{-8\lambda} = \lambda^3 e^{-24\lambda}.$$

(d) The posterior for λ is proportional to the likelihood $\times$ prior,

$$\lambda^3 e^{-24\lambda} \times \lambda^{-1/2} = \lambda^{5/2} e^{-24\lambda} = \lambda^{7/2-1} e^{-24\lambda},$$

which can be recognized as Gamma $\mathcal{G}a(7/2, 24)$ distribution. Since the mean of $\mathcal{G}a(\alpha, \beta)$ is α/β, the mean of the posterior is

$$\hat{\lambda}_B = \frac{7/2}{24} = 0.1458,$$

which is the Bayes estimator, in this case.

10.16. Aerobic Capacity. The peak oxygen intake per unit of body weight, called *the aerobic capacity* of an individual performing a strenuous activity is a measure of work capacity. For comparative study, measurements of aerobic capacities are recorded (Frisancho, 1975) for a group of 20 Peruvian Highland natives and for a group of 10 Peruvian lowlanders acclimatized as adults to high altitudes. The measurements are taken on a bicycle ergometer at high altitude ($ml\ kg^{-1}\ min^{-1}$).

	Peruvian Highland Natives	Peruvian Lowlanders Acclimatized as Adults
Sample mean	46.3	38.0
Sample st. deviation	5.0	5.2

(a) Test the hypothesis that the population mean aerobic capacities are the same against the one sided alternative. Assume equality of population variances and take $\alpha = 0.05$.

(b) If you are to repeat this experiment, what sample size (per group) will give you the power of 90% to detect the difference between the means of magnitude 4, if you adopt that the common population variance is $\sigma^2 = 5^2 = 25$. The level of the test, α is to be kept at 5%.

```
%% aerobic

x1bar = 46.3;
x2bar = 38.0;
s1 = 5; s2 = 5.2;
n1 = 20; n2 = 10;
%
sp = sqrt( ((n1 -1 )*s1^2 + (n2-1)*s2^2 )/(n1 + n2 - 2))
% sp = 5.0651
t = (x1bar - x2bar)/(sp * sqrt(1/n1 + 1/n2))
% t = 4.23

%RR (tinv(0.95, 28), +infinity
tinv(0.95, n1 + n2 - 2) %1.7011
 % RR = [1.7011, +infinity), Reject H0 since t is in RR
%pvalue
pval = 1 - tcdf(t, n1 + n2 - 2)
%pbval = 1.1278e-04 < 0.05 rehject H0


%(b)

sigma = 5; alpha = 0.05; beta = 0.1; delta = 4;

n = 2 * sigma^2/delta^2 * (norminv(1-alpha) + norminv(1-beta))^2
%n = 26.7620, take n = 27
```

Question [value: 1 point each]	True	False
1. The mean of a random variable is always larger than the median if its distribution is skewed left.		$\otimes$
2. One can always find a diagnostic test that can achieve specificity of 100%.	$\otimes$	
3. For any three events A,B, and C, probability of $A \cup B$ is always larger or equal to the probability of $B \cap C$.	$\otimes$	
4. Significance levels of a test α is set to be one of 1%, 5%, or 10%, by National Institute of Health (NIH) bylaws.		$\otimes$
5. The first and third quartiles Q_1 and Q_3 are 0.25 and 0.75 quantiles (equivalently) 25% and 75% percentages of a sample. The interquartile range is $Q_3 - Q_1$ represents a measure of variability in the sample.	$\otimes$	
6. Bayesian probability of H_0 and classical p value always coincide. It is only the philosophical standpoint that separates the two.		$\otimes$
7. Chi-square χ^2 random variable with n degrees of freedom is distributed as the sum of squares of n independent standard normals.	$\otimes$	
8. Central limit theorem allows to use normal distribution to approximate binomial probabilities when n is large. This approximation is good unless p is small, when Poisson approximation is better.	$\otimes$	
9. To find the power of a test, $1-\beta$ one needs to know the sample size, and the magnitude of discrepancy from H_0, but not the level α, since α and β are independent.		$\otimes$
10. In Bayesian calculation, a binomial likelihood and gamma prior form a conjugate pair from which it is easy to find the posterior.		$\otimes$
11. The power of a test conceptually represents the probability that true H_0 will not be rejected.		$\otimes$
12. A rejection region for testing a normal mean against the one sided alternative, when variance σ^2 is known, was found to be $[1.96, \infty)$. If the test statistic Z was found to be 2.2, the H_0 is not rejected since the p-value is $2.20 - 1.96 = 0.24 > 5\%$.		$\otimes$
13. The 45th percentile of a standard normal distribution is negative.	$\otimes$	
14. The total length of Wald's 95% confidence interval based on $n = 100$ trials with $X = 50$ successes is greater than 0.1.	$\otimes$	
15. If a single measurement is normal with mean 5 and variance 2, then the average of 10 such measurements would have mean 5 and variance $\frac{2}{10}$.	$\otimes$	
16. A confidence interval for normal population variance σ^2 involves quantiles of t distribution and sample variance s^2.		$\otimes$
17. Bayes rule finds the probability of a hypothesis if an event occurred, that is, changes a postulated prior probability to the posterior probability. This change sometimes could be drastic, when experimental evidence is strong. For example, if before the experiment one believes that probability of hypothesis is $p = 0.000001$, then after the experiment it could be possible to infer that $p = 0.999999$.	$\otimes$	
18. The probabilities of type I and II errors, α and β, can be thought as probabilities of rejection region when H_0 is false, and of acceptance region when H_1 is false, respectively.		$\otimes$
19. Point estimators of parameters are random variables with specific sampling distributions. If this distribution is continuous, the probability that estimator matches the parameter it estimates is 0, that is, the parameter in continuous distributions can **never** be estimated exactly.	$\otimes$	
20. Suppose we are independently testing that 1000 hypotheses concerning whether 1000 genes are over or under expressed at level 5% each. Suppose that the truth is that none of the genes is over/under expressed. We will still find about 50 genes that are significantly expressed, so called "false discoveries."	$\otimes$	