BST 140.751
Problem Set 2

1 Distributions

1. In a random sample of 100 subjects with low back pain, 27 reported an improvement in symptoms after execise therapy.

   A. Give and interpret (in the comments) an interval estimate for the true proportion of subjects who respond to exercise therapy.

   B. Plot the likelihood for the true proportion of subjects with low back pain.

   C. Plot the posterior and give equi-tail and HPD credible intervals.

   D. Give the likelihood interval.

2. Let p denote the unknown proportion of rocks in a riverbed that are sedimentary in type. Suppose that $X = 12$ of a sample of $n = 20$ rocks collected in random locations are found to be sedimentary in type.

   A. Plot the likelihood for the parameter p and interpret.

   B. From your graphs, determine the value of $\hat{p}$ of p where the curve reaches its maximum. Does this value for the maximum make intuitive sense? Comment in one or two sentences.

   C. Show that the point that maximizes the binomial likelihood is always X/n.

   D. Use the CLT to create a confidence interval for the true proportion of rocks that are sedimentary. Interpret your results.

   E. A much larger study is planned and the researchers would like to know how large n should be to have a margin of error on the estimate for the proportion of sedimentary rocks that is no larger than $.01$ for a 95% confidence interval? Use the fact that $p(1-p) \leq 1/4$. Also try the calculation with the estimate of p from the current study.

3. Using a computer, generate 1000 Binomial random variables for $n = 10$ and $p = .3$ Calculate the percentage of times that

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$$

   contains the true value of p. Here $\hat{p} = X/n$ where X is each binomial variable. Do the intervals appear to have the coverage that they are supposed to?

4. Repeat the calculation only now use the interval

$$\tilde{p} \pm 1.96\sqrt{\tilde{p}(1-\tilde{p})/n}$$

   where $\tilde{p} = (X+2)/(n+4)$. Does the coverage appear to be closer to .95?

5. Repeat this comparison (parts a. - d.) for $p = .1$ and $p = .5$. Which of the two intervals appears to perform better?

6. A statistic is called "sufficient" if the likelihood only depends on that statistic. That is only if the likelihood of a parameter given the full data is proportional to a likelihood only depending on the sufficient statistics.

   A. Let $X_1, \ldots, X_n$ be iid Bernoulli(p). What is the sufficient statistic for p?

   B. Let $X_1, \ldots, X_n$ be iid Poisson(μ). What is the sufficient statistics for μ?

   C. Let $X_1, \ldots, X_n$ be iid Normal(μ, σ^2). What is the sufficient statistics for (μ, σ^2).

   D. Let $X_1, \ldots, X_n$ be Uniform(θ, $\theta + 1$). What is the sufficient statistic for θ?

   E. Let $X_1, \ldots, X_n$ be Gamma(α, β). What are the sufficient statistics for α and β?

7. Let $X_1, \ldots, X_n$ be iid.

   A. If $X_i \sim$Bernoulli(p) then $Y = \sum_{i=1}^n$ is Binomial. Argue that the likelihoods using $\{X_i\}_{i=1}^n$ and Y are equivalent. Do the MLEs agree?

   B. If $X_i \sim$Poisson(μ) then $Y = \sum_{i=1}^n X_i$ is Poisson($n\mu$). Argue that the likelihoods using $\{X_i\}_{i=1}^n$ and Y are equivalent. Do the MLEs agree?

   C. Why do the likelihoods and MLEs agree?

8. Consider let Y_i ($i = 1, \ldots, I$) be iid discrete random variables that take the values $1, \ldots, K$ with probabilities p_k so that $0 \leq p_k \leq 1$ and $\sum_{k=1}^K p_k = 1$. Let $X_{ik} = 1$ if $Y = k$ and 0 otherwise. Let $X_i = (X_{i1}, \ldots, X_{iK})$. X_i is a multivariate Bernoulli with the Bernoulli distribution being the special case when $K = 2$.

   A. Write out the likelihood for $(p_1, \ldots, p_k)$.

   B. Argue that the sufficient statistics are $n_k = \sum_{i=1} X_{ik}$.

   C. Show that the ML estimate of p_k is n_k/n.

   D. Derive the mass function for $n = (n_1, \ldots, n_k)$? (Hint: look it up; it's called the multinomial distribution.)

   E. Argue that the any collapsed subset is also multinomial. For example $(n_1 + n_2, n_3, \ldots, n_k)$ is multinomial with probabilities $p_1 + p_2, p_3, \ldots, p_k$.

9. Let $X_1 \ldots X_k$ be independent so that $X_k \sim$ Poisson(μ_k). Argue That $(X_1, \ldots, X_k) \mid N = \sum_{k=1}^K X_k$ is multinomial with sample size N and probabilities $p_k = \mu_k / \sum_{l=1}^K \mu_l$.

10. A profile likelihood for a two parameter likelihood, say $\mathcal{L}(\mu, \theta)$ is the function $PL(\theta) = \mathcal{L}(\hat{\mu}(\theta), \theta)$ where $\hat{\mu}(\theta)$ is the ML estimate for μ with θ held fixed as if it were known. Conversely, the profile likelihood for μ is $PL(\mu) = \mathcal{L}(\mu, \hat{\theta}(\mu))$ where $\hat{\theta}(\mu)$ is the ML estimate for θ with μ held fixed as if it were known.

    A. Let $X_1, \ldots, X_N$ be iid N(μ, θ). Calculate the profile likelihoods for θ and μ.

    B. Argue why it's called a "profile" likelihood.

    C. Argue that argmax$_\mu PL(\mu)$ is the ML estimate for μ. (That is, the maximum of the profile likelihood is the maximum of the overall likelihod.)

11. A large survey of over 100,000 births in South Wales during the period 1956-1962 gave an incidence rate for spina bifida of 4.12 per 1,000 births. In a random sample of 1000 births, compute the probability of observing (i) no cases, (ii) one case, (iii) two cases. Using the following two approaches

    A. The exact distribution based on the binomial distribution

    B. Approximate probabilities based on a Poisson approximation to the binomial

2 Bayesian statistics

1. Let $X_1, \ldots, X_n \mid p$ be Bernoulli(p) and $p \sim Beta(\alpha, \beta)$. Derive the posterior distribution and give its mean and variance.

2. Let $X_1, \ldots, X_n \mid \lambda$ be Poisson$(t_i\lambda)$ (respectively) and $\lambda \sim Gamma(\alpha, \beta)$. Derive the posterior distribution and give its mean and variance.

3. Let $n = (n_1, \ldots, n_k)|\theta = (\theta_1, \ldots, \theta_k)$ be multinomial(N, θ) and $\theta \sim Dirichlet(\alpha)$ where $\alpha = (\alpha_1, \ldots, \alpha_k)$. Derive the posterior for $theta$ and calculate its mean.

4. Let $X_1, \ldots, X_n \mid \mu, \sigma^2$ be iid Normal(μ, σ^2) and $\mu \mid \sigma^2 \sim Normal(\mu_0, \sigma^2\tau)$ and $\sigma^{-2} \sim Gamma(\alpha, \beta)$. Derive the posterior for (μ, σ).

3 The Delta method

1. Assume the fact that if $X \sim$Poisson(λt) then $\frac{X-\lambda t}{\sqrt{t\lambda}} \to N(0, 1)$ for large t.

    A. Use the delta method to calculate an interval for $\log(\lambda)$.

    B. Suppose that $Y \sim$ Poisson$(\lambda_2 t_2)$. Create a confidence interval for $\log(\lambda/\lambda_2)$.

    C. Let $X \sim$Binomial(n, p). Use the delta method to get an interval estimate for $\sqrt{p}$.

2. An example of the multivariate delta method says that if $\bar{X}$ is a random p dimensional random vector so that $\sqrt{n}(\bar{X} - \mu) \to N(0, \Sigma)$, f is a function from $\mathbb{R}^p \to \mathbb{R}^1$ and$\nabla f(\bar{X})$ is the gradient vector of f evaluated at $\bar{X}$ then $\sqrt{n}\{f(\bar{X}) - f(\mu)\} \to N\{0, \nabla f(\bar{X})'\Sigma\nabla f(\bar{X})\}$ (in distribution).

    A. Use the multivariate delta method to derive the standard formula for the CI for the relative risk.

    B. Use the multivariate delta method to derive the standard formula for the CI for the odds ratio.

4 Multivariate means, variances and normals

1. Let X be a multivariate vector with mean μ. Show that $E[AX + b] = A\mu + b$.
2. Consider the previous problem; assume that Var$(X) = \Sigma$. Show that Var$(AX + b) = A\Sigma A'$.
3. Show that $E[(X - \mu)(X - \mu)'] = E[XX'] - \mu\mu'$.
4. Argue that Var(X) is non-negative definite.
5. Let $C(X, Y)$ be the multivariate covariance function, $E[(X - \mu_x)(Y - \mu_y)']$. Show that $C(X, Y) = E[XY'] - \mu_X\mu_y'$.
6. Show that $C(X_1 + X_2, Y) = C(X_1, Y) + C(X_2, Y)$.
7. Argue that $C(X, Y) = C(Y, X)'$.
8. Argue that $Var(X + Y) = Var(X) + C(X, Y) + C(Y, X) + V(Y)$.
9. Argue that $C(AX, BY) = AC(X, Y)B'$.
10. Let $X \sim N(0, I)$. Argue that $aX/\sqrt{a'a} \sim N(0, 1)$ for any non-zero vector a.
11. Let $X \sim N(0, I)$. Argue that if $AA' = I$ then $AX \sim N(0, I)$. Argue geometrically why this occurs.
12. Let X_i for $i = 1, \ldots, I$ be iid k dimensional vectors from a distribution with mean μ and variance Σ. What is the mean and variance of the multivariate pointwise sample average of the vectors?
13. Let X_i be iid k dimensional vectors from a distribution with mean μ and variance Σ. Give an unbiased estimate of Σ when μ is known.
14. Consider a covariance matrix that is of the form

    $$\sigma^2\mathbf{I} + \theta\mathbf{1}\mathbf{1}'$$

    where σ^2 and θ are positive constants and $\mathbf{1}$ is a vector of ones. Argue that this matrix describes random vectors where every pair of elements of the vector are equally correlated and every elemant has the same variance. Give this correlation and variance.
15. Let $X = (X_1' \; X_2')' \sim N(\mu, \Sigma)$.
    A. Derive the marginal distribution of X_1
    B. Derive the conditional distribution $X_1 \mid X_2$.
16. Let $X \mid \mu \sim N(\mu, \Sigma)$ and $\mu \sim N(\alpha, \tau I)$. Derive the distribution of $\mu \mid X$.
17. Argue that if $Y \sim N(\mu, \Sigma)$, the quadratic form $(Y - \mu)'\Sigma^{-1}(Y - \mu)$ is χ^2_p.

5 Linear models

1. Let $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$.

   A. Derive the MLEs for β_0, β_1 and σ^2.

   B. Relate β_1 to the correlation between Y_i and X_i.

   C. Suppose that you standardize (i.e. take $(Y_i - \bar{Y})/S_y$ and $(X_i - \bar{X})/S_X$) X_i and Y_i. Derive the estimates of β_0 and β_1.

2. Let $Y_{ij} = \alpha_0 + \beta_j + \epsilon_{ij}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

   A. Write out the design matrix for the associated linear model.

   B. Show what the estimates are under the following constraints:

      i. $\alpha_0 = 0$
      ii. $\beta_1 = 0$
      iii. $\beta_J = 0$
      iv. $\sum_{j=1}^{J} \beta_j = 0$

3. Let Σ be a known matrix. Consider the model $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, \Sigma)$. Derive the ML estimate of β.

4. Let P be a rotation matrix and consider the model $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, \sigma^2 I)$. Suppose someone gave you the ML estimates for $\tilde{\beta}$ and $\tilde{\sigma}^2$ from fitting the model $\tilde{Y} = \tilde{X}\tilde{\beta} + \tilde{\epsilon}$ where $\tilde{Y} = PY$ and $\tilde{X} = PX$ and $\tilde{\epsilon} \sim N(0, \tilde{\sigma}^2)$. Relate these estimates to the ML estimates of β and σ^2.

5. Let $Y \mid \beta \sim N(X\beta, \sigma^2 I)$ and $\beta \sim N(\beta 0, \tau^2 I)$. What is the posterior distribution of β?

6. Consider the model $Y = X\beta + \epsilon$. Let F be an invertible $p \times p$ matrix and $\tilde{X} = XF$.

   A. Consider another model $Y = \tilde{X}\tilde{\beta} + \epsilon$. Argue that the models are equivalent.

   B. Show that the least squares estimate of $\tilde{\beta}$ from the second model is $F^{-1}\hat{\beta}$ where $\hat{\beta}$ is the least squares estimate from the first model.

   C. Suppose that you have a linear regression equation where one of the regressors is temperature. Use the results above to relate the beta coefficients if the regressor is input as Celsius or Fahrenheit.

7. Consider a linear model with iid errors $N(0, \sigma^2)$ errors. Show that $\frac{1}{n-p} e'e$, where e is the vector of residuals, is the ML estimate of σ^2. Further show that this estimate is unbiased.

   A. Argue that $\frac{1}{\sigma^2}(y - X\beta)'(y - X\beta)$ is χ^2_n

   B. Argue that $\frac{1}{\sigma^2} e'e$ is χ^2_{n-p}.

   C. Argue that $\frac{1}{\sigma^2}(y - X\beta)'X(X'X)^{-1}X'(y - X\beta)$ is χ^2_p.

   D. In each of the above cases, use the expected value calculation for quadratic forms to verify that the expected values equals the Chi squared df.

6 Computing and analysis

1. Write an R function that takes a Y vector and X matrix and obtains the least squares fit for the associated linear model.

2. Write and R function that takes an $n \times p$ data matrix, X, and "whitens" it via subtracting out a mean and multiplying by a matrix so that the resulting matrix has (p) sample column means of 0 and $p \times p$ sample covariance matrix of I.

3. You collect 100 blood pressure measurements from a population. Twenty one exhibit high blood pressure. Using a beta prior, plot the posteriors assuming: Beta(1, 1), Beta(2, 2), Beta(.5, .5). Calculate 95% equi-tail and HPD credible intervals.

4. You now collect 100 measurements from an otherwise similar population with different diets; 15 have high blood pressure. Plot the posteriors for the relative risk, risk difference and risk ratio. Assume Beta(1,1) priors for both populations.

5. One nuclear reactor test failed 51 times after having been monitored for $1,000$ days. Assuming a Gamma(.1,.1) prior and a Poisson model, plot the posterior for the failure rate. Give a credible interval (HPD and equi-tail).

6. A second reactor failed 25 times for 600 monitoring days. Plot the posterior and give credible intervals (HPD and equi-tail) for the relative rate with the other reactor.