*6. The beta distribution with parameters $\alpha > 0$ and $\beta > 0$ has density

$$f_{\alpha,\beta}(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & x \in (0,1); \\ 0, & \text{otherwise.} \end{cases}$$

Suppose $X_1, \ldots, X_n$ are i.i.d. from a beta distribution.

a) Determine a minimal sufficient statistic (for the family of joint distributions) if α and β vary freely.

b) Determine a minimal sufficient statistic if $\alpha = 2\beta$.

c) Determine a minimal sufficient statistic if $\alpha = \beta^2$.

---

a) The joint densities are

$$\exp\left[(\alpha-1)\sum_{i=1}^{n}\log x_i + (\beta-1)\sum_{i=1}^{n}\log(1-x_i) + n\log\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right],$$

a full rank exponential family with

$$T = \left(\sum_{i=1}^{n}\log X_i, \sum_{i=1}^{n}\log(1-X_i)\right)$$

a minimal sufficient statistic.

b) Now the joint densities are

$$\exp\left[(\beta-1)\sum_{i=1}^{n}\log\left[x_i^2(1-x_i)\right] + \sum_{i=1}^{n}\log x_i + n\log\frac{\Gamma(3\beta)}{\Gamma(2\beta)\Gamma(\beta)}\right],$$

a full rank exponential family with minimal sufficient statistic

$$\sum_{i=1}^{n}\log\left[x_i^2(1-x_i)\right] = 2T_1 + T_2.$$

c) The densities, parameterized by β, are

$$p_\beta(x) = \exp\left[(\beta^2 - 1)T_1(x) + (\beta - 1)T_2(x) + n \log \frac{\Gamma(\beta + \beta^2)}{\Gamma(\beta)\Gamma(\beta^2)}\right].$$

Suppose $p_\beta(x) \propto_\beta p_\beta(y)$. Then

$$\frac{p_2(x)}{p_1(x)} = \frac{p_2(y)}{p_1(y)} \text{ and } \frac{p_3(x)}{p_1(x)} = \frac{p_3(y)}{p_1(y)}.$$

Taking the logarithm of these and using the formula for p_β,

$$3T_1(x) + T_2(x) + n \log 20 = 3T_1(y) + T_2(y) + n \log 20,$$

and

$$8T_1(x) + 2T_2(x) + n \log 495 = 8T_1(y) + 2T_2(y) + n \log 495.$$

These equations imply $T(x) = T(y)$, and T is minimal sufficient by Theorem 3.11.

*7. *Logistic regression.* Let $X_1, \dots, X_n$ be independent Bernoulli variables, with $p_i = P(X_i = 1)$, $i = 1, \dots, n$. Let $t_1, \dots, t_n$ be a sequence of known constants that are related to the p_i via

$$\log \frac{p_i}{1 - p_i} = \alpha + \beta t_i,$$

where α and β are unknown parameters. Determine a minimal sufficient statistic for the family of joint distributions.

The statistic $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n t_i X_i\right)$, is minimal sufficient.

*8. The multinomial distribution, derived later in Section 5.3, is a discrete distribution with mass function

$$\frac{n!}{x_1! \times \cdots \times x_s!} p_1^{x_1} \times \cdots \times p_s^{x_s},$$

where $x_0, \ldots, x_s$ are nonnegative integers summing to n, where $p_1, \ldots, p_s$ are nonnegative probabilities summing to one, and n is the sample size. Let $N_{11}, N_{12}, N_{21}, N_{22}$ have a multinomial distribution with n trials and success probabilities p_{11}, p_{12}, p_{21}, p_{22}. (A common model for a two-by-two contingency table.)

a) Give a minimal sufficient statistic if the success probabilities vary freely over the unit simplex in $\mathbb{R}^4$. (The unit simplex in $\mathbb{R}^p$ is the set of all vectors with nonnegative entries summing to one.)

b) Give a minimal sufficient statistic if the success probabilities are constrained so that $p_{11}p_{22} = p_{12}p_{21}$.

a) The statistic (N_{11}, N_{12}, N_{21}) is minimal sufficient. (The statistic (N_{11}, N_{12}, N_{21}, N_{22}) is also minimal sufficient.)

b) With the constraint, $(N_{11} + N_{12}, N_{11} + N_{21})$ is minimal sufficient.

*9. Let f be a positive integrable function on $(0, \infty)$. Define

$$c(\theta) = 1 \Big/ \int_\theta^\infty f(x)\,dx,$$

and take $p_\theta(x) = c(\theta)f(x)$ for $x > \theta$, and $p_\theta(x) = 0$ for $x \le \theta$. Let $X_1, \dots, X_n$ be i.i.d. with common density p_θ.

a) Show that $M = \min\{X_1, \dots, X_n\}$ is sufficient.

b) Show that M is minimal sufficient.

---

a) The joint density is zero unless $x_i > \theta$, $i = 1, \dots, n$, that is, unless $M(x) = \min\{x_1, \dots, x_n\} > \theta$. Using this, the joint density can be written as

$$p_\theta(x) = c^n(\theta) I\{M(x) > \theta\} \prod_{i=1}^n f(x_i),$$

and $M(X)$ is sufficient by the factorization theorem.
b) If $p_\theta(x) \propto_\theta p_\theta(y)$, then the region where the two functions are zero must agree, and $M(x)$ must equal $M(y)$. So M is minimal sufficient by Theorem 3.11.

*17. Let $X, X_1, X_2, \dots$ be i.i.d. from an exponential distribution with failure rate λ (introduced in Problem 1.30).

a) Find the density of $Y = \lambda X$.

b) Let $\overline{X} = (X_1 + \cdots + X_n)/n$. Show that $\overline{X}$ and $(X_1^2 + \cdots + X_n^2)/\overline{X}^2$ are independent.

---

a) If $y > 0$, then $P(Y \le y) = P(\lambda X \le y) = \int_0^{y/\lambda} \lambda e^{-\lambda x}\,dx = 1 - e^{-y}$. So Y has density $d(1 - e^{-y})/dy = e^{-y}$, $y > 0$, the standard exponential density.
b) The joint densities are $\lambda^n \exp\{-n\lambda\overline{x}\}$, a full rank exponential family with $T = \overline{X}$ a complete sufficient statistic. Let $Y_i = \lambda X_i$, so that regardless of the value of λ, $Y_1, \dots, Y_n$ are i.i.d. from the standard exponential distribution. Then $(X_1^2 + \cdots + X_n^2)/\overline{X}^2 = (Y_1^2 + \cdots + Y_n^2)/\overline{Y}^2$ is ancillary, and independence follows by Basu's theorem.