1. The binomial distribution as an exponential family

   (a) For $\theta = \log\{p/(1-p)\}$ we get

   $$p = \frac{e^\theta}{1+e^\theta}, \quad 1-p = \frac{1}{1+e^\theta}$$

   so we can write

   $$f(x;\theta) = \binom{n}{x} \frac{e^{\theta x}}{(1+e^\theta)^x} \frac{1}{(1+e^\theta)^{n-x}} = \binom{n}{x} \frac{e^{\theta x}}{(1+e^\theta)^n} = \binom{n}{x} e^{\theta x - c(\theta)}$$

   which identifies the family as a canonical exponential family with canonical statistic $t(x) = x$ and

   (b) $c(\theta) = n \log(1+e^\theta)$;

   (c) We get

   $$\mathbf{E}(X) = \tau(\theta) = c'(\theta) = \frac{e^\theta}{1+e^\theta}$$

   and

   $$\mathbf{V}(X) = c''(\theta) = n\frac{e^\theta(1+e^\theta) - (e^\theta)^2}{(1+e^\theta)^2} = n\frac{e^\theta}{(1+e^\theta)^2} = np(1-p).$$

   (d) $\eta = \tau(\theta) = np$;

   (e) The likelihood equation thus becomes

   $$\hat\eta = n\hat p = x$$

   which has the familiar solution $\hat p = x/n$.

2. Estimation of risk.

   (a) For $\theta = \log\lambda$ we get $\lambda = e^\theta$ and

   $$f(x;\theta) = \prod_i \frac{(\lambda N_i)^{x_i}}{x_i!} e^{-\lambda N i} = \left\{\prod_i \frac{N_i^{x_i}}{x_i!}\right\} e^{\theta \sum_i x_i - e^\theta \sum_i N i}$$

   which identifies the family as canonical exponential with

   (b) canonical sufficient statistic $t(x) = \sum_i X_i$ and

   $$c(\theta) = e^\theta \sum_i N_i$$

   so the mean of the sufficient statistic is

   $$\mathbf{E}(\sum_i X_i) = \tau(\theta) = c'(\theta) = e^\theta \sum_i N_i$$

and the variance is

$$\mathbf{V}(\sum_i X_i) = c''(\theta) = e^\theta \sum_i N_i.$$

This could also have been found by more traditional methods!

(c) The mean value parameter is

$$\eta = \tau(\theta) = e^\theta \sum_i N_i = \lambda \sum_i N_i.$$

(d) The maximum likelihood estimate $\hat{\lambda}$ of λ is determined by equating the canonical sufficient statistic to its expectation, i.e.

$$\sum_i X_i = \hat{\lambda} \sum_i N_i$$

so

$$\hat{\lambda} = \frac{\sum_i X_i}{\sum_i N_i}$$

i.e. the ratio between the total number of occurrences and the total number at risk.

(e) The Fisher information for λ is found as minus the expectation of

$$\frac{\partial^2}{\partial \lambda^2}\left\{(\log \lambda)\sum_i X_i - \lambda \sum_i N_i\right\}$$

to be

$$i(\lambda) = \frac{\sum_i N_i}{\lambda}.$$

Since λ is a linear function of the mean value parameter, the maximum likelihood estimate is efficient and thus

$$\mathbf{V}(\hat{\lambda}) = i(\lambda)^{-1} = \frac{\lambda}{\sum_i N_i},$$

so to obtain a small variance it is important to study a large number of individuals at risk.

3. Hardy–Weinberg equilibrium.

(a) Inserting μ into the probability mass function yields

$$f(x;\mu) = \binom{n}{x_{AA}, x_{Aa}, x_{aa}} \mu^{2x_{AA}} (2\mu)^{x_{Aa}} (1-\mu)^{x_{Aa}} (1-\mu)^{2x_{aa}}.$$

Letting

$$\theta = \log\frac{\mu}{1-\mu}, \quad \mu = \frac{e^\theta}{1+e^\theta}, \quad 1-\mu = \frac{1}{1+e^\theta}$$

and rearranging, using $n = x_{AA} + x_{Aa} + x_{aa}$ yields

$$f(x;\mu) = \binom{n}{x_{AA}, x_{Aa}, x_{aa}} 2^{x_{Aa}}\ e^{\theta(2x_{AA}+x_{Aa})} \frac{1}{(1+e^\theta)^{2n}}$$

which identifies an exponential family with canonical statistic $t(X) = 2X_{AA} + X_{Aa}$;

(b) and canonical parameter

$$\theta = \log \frac{\mu}{1-\mu}.$$

(c) The log-normalizing function is

$$c(\theta) = 2n \log(1 + e^\theta).$$

(d) The mean value parameter is

$$\tau(\theta) = \mathbf{E}(2X_{AA} + X_{Aa}) = 2n\mu.$$

(e) So the likelihood equation becomes

$$2x_{AA} + x_{Aa} = 2n\mu$$

which yields the MLE as

$$\hat{\mu} = \frac{2x_{AA} + x_{Aa}}{2n}$$

the total number of A-alleles divided by the total number of alleles.

4. Non-regular exponential families.

(a) The probability densities

$$f(x;\theta) = \frac{e^{\theta x - c(\theta)}}{1+x^4}, \quad \text{for } x \geq 0,$$

form an exponential family for $\theta \leq 0$ with log-normalising function

$$c(\theta) = \log\left(\int_0^\infty \frac{e^{\theta x}}{1+x^4} dx\right).$$

The canonical sufficient statistic is x and since this is an exponential family the expectation and variance of X are

$$\mathbf{E}_\theta(X) = c'(\theta) \quad \text{and} \quad \mathbf{V}_\theta(X) = c''(\theta), \quad \text{for } \theta < 0. \qquad (*)$$

(b) By direct integration

$$\mathbf{E}_0(X) = \int_0^\infty \frac{xe^{-c(0)}}{1+x^4}\,dx = \frac{\pi}{4}e^{-c(0)} = 1/\sqrt{2},$$

since

$$e^{c(0)} = \int_0^\infty \frac{1}{1+x^4}\,dx = \frac{\pi}{4}\sqrt{2}.$$

From $(*)$ it follows that $\mathbf{E}_\theta(X)$ is strictly increasing since its derivative $c''(\theta)$ is the variance which is positive. It follows that $\mathbf{E}_\theta(X)$ cannot exceed $1/\sqrt{2}$, so that the likelihood equation has no solution when $x > 1/\sqrt{2}$.

(c) The derivative of $\log f(x;\theta)$ with respect to θ is $x - c'(\theta)$ which is always positive when $x > 1/\sqrt{2}$ and the likelihood function is therefore itself increasing. It follows that $f(x;\theta)$ has its maximum at $\theta = 0$ for these cases, hence that $\hat{\theta} = 0$ if $x \geq 1/\sqrt{2}$.

(d) Indeed g integrable so we find as above that $\Theta = (-\infty, 0]$. In fact

$$e^{c(0)} = \int_0^\infty \frac{1}{1+x^2}\,dx = \frac{\pi}{2}.$$

(e) But here we get

$$\mathbf{E}_0(X) = \int_0^\infty \frac{xe^{-c(0)}}{1+x^2}\,dx = \infty,$$

so that $\tau(\theta) = \mathbf{E}_\theta(X)$ is not bounded. As $\tau(\theta)$ is continuous and $\tau(-\infty) = 0$, the likelihood equation has a (unique) solution for all $0 < x < \infty$.