9.3 a) $\mathbb{E}\hat{\vartheta}_1 = \mathbb{E}\bar{Y} - \frac{1}{2} = \frac{(\vartheta+1+\vartheta)}{2} - \frac{1}{2} = \frac{2\vartheta+1}{2} - \frac{1}{2} = \vartheta$. Unbiased.

$\mathbb{E}\hat{\vartheta}_2$ needs $\mathbb{E}Y_{(n)}$.

$$f(y) = n F(y)^{n-1} f(y) = n(y-\vartheta)^{n-1}\cdot 1$$
$$\mathbb{E}Y_{(n)} = \int_{\vartheta}^{\vartheta+1} y\cdot n\cdot(y-\vartheta)^{n-1}\,dy = n\int_0^1 (u+\vartheta)u^{n-1}\,du = n\int_0^1 u^n\,du + n\vartheta\int_0^1 u^{n-1}\,du =$$
$$= n\left(\frac{u^{n+1}}{n+1}\right)\Big|_0^1 + n\vartheta\left(\frac{u^n}{n}\right)\Big|_0^1 = \frac{n}{n+1} + \vartheta$$ So unbiased.

b) $\mathbb{V}\hat{\vartheta}_1 = \mathbb{V}\bar{Y} = \frac{\mathbb{V}Y}{n} = \frac{1}{12n}$

For $\mathbb{V}\hat{\vartheta}_2$, notice that variance is translation invariant: we might as well set $\vartheta=0$ for the variance calculation. So we need $\mathbb{V}Y_{(n)}$ where $Y_i \sim \text{Uniform}(0,1)$.
We calculate:

$$\mathbb{E}Y_{(n)} = \frac{n}{n+1} \quad \text{so} \quad (\mathbb{E}Y_{(n)})^2 = \frac{n^2}{(n+1)^2}$$

$$\mathbb{E}(Y_{(n)}^2) = \int_0^1 n y^{n+1}\,dy = n/n+2$$

So $\mathbb{V}\hat{\vartheta}_2 = \mathbb{V}Y_{(n)} = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n(n+1)^2 - n^2(n+2)}{(n+1)^2(n+2)}$

The relative efficiency is $\frac{n(n+1)^2 - n^2(n+2)}{(n+1)^2(n+2)} \cdot 12n =$

$$= \frac{12n^2(n^2+2n+1-n^2-2n)}{(n+1)^2(n+2)} = \frac{12n^2}{(n+1)^2(n+2)}$$

9.15

Claim: Both $\hat{\vartheta}_1 = \bar{Y} - \frac{1}{2}$ and $\hat{\vartheta}_2 = Y_{(n)} - \frac{n}{n+1}$ are consistent.

$\hat{\vartheta}_1$: $\lim \mathbb{V}\hat{\vartheta}_1 = \lim \frac{12}{n} = 0.$ $\blacksquare$

$\hat{\vartheta}_2$: $\lim \mathbb{V}\hat{\vartheta}_2 = \lim \frac{12n^2}{(n+1)^2(n+2)} = 0$ the denominator has higher degree. $\blacksquare$

9.27

For small enough ε, we need

$$\mathbb{P}(\vartheta-\varepsilon \le Y_{(1)} \le \vartheta+\varepsilon) = F_{(1)}(\vartheta+\varepsilon) - F_{(1)}(\vartheta-\varepsilon) =$$
$$1-\left(1-\left(1-\frac{\vartheta-\varepsilon}{\vartheta}\right)^n\right) = 1-1+\left(1-\frac{\vartheta-\varepsilon}{\vartheta}\right)^n = \left(\frac{\vartheta-\vartheta+\varepsilon}{\vartheta}\right)^n = \left(\frac{\varepsilon}{\vartheta}\right)^n$$

$\lim_{n\to\infty} \left(\frac{\varepsilon}{\vartheta}\right)^n = 0.$ Hence inconsistent.

We know by now:

- How to evaluate estimators (bias, variance, consistency, sufficiency)
- How to improve estimators (Rao-Blackwell)

We don't yet know how to find estimators to begin with.

Method of Moments

Recall the k^{th} moment is $\mu_k' = \mathbb{E}[Y^k]$.

The k^{th} sample moment is $m_k' = \frac{1}{n} \sum Y_i^k$.

Pick estimators by solving $\mu_k' = m_k'$ for the parameter of interest.

Example

$Y_i \sim \text{Uniform}(0, \vartheta)$.

$\mu_1' = \mathbb{E}Y = \vartheta/2$. $\quad m_1' = \bar{Y}$

So we set $\frac{\vartheta}{2} = \bar{Y}$ and get the estimator $\hat{\vartheta} = 2\bar{Y}$.

Method of Moments estimators will tend to be consistent, but rarely MVUE.

In this example, for instance, $Y_{(n)}$ emerges from the factorization criterion and has lower variance.

__Example__ $Y_1, \ldots, Y_n \sim \text{Gamma}(\alpha, \beta)$.

We know about the Gamma distribution that

$$\mu_1' = \mu = \alpha\beta \qquad \mu_2' = \sigma^2 + \mu^2 = \alpha\beta^2 + \alpha^2\beta^2$$

So we get a system of equations

$$\frac{1}{n}\sum Y_i = \bar{Y} = \alpha\beta \qquad \text{so} \quad \beta = \bar{Y}/\alpha$$

$$Y^{(2)} = \frac{1}{n}\sum Y_i^2 = \alpha\beta^2 + \alpha^2\beta^2 = (\alpha+\alpha^2)\beta^2 = (\alpha+\alpha^2)\frac{\bar{Y}^2}{\alpha^2} = \frac{1+\alpha}{\alpha}\bar{Y}^2$$

So $\alpha Y^{(2)} = \bar{Y}^2 + \alpha\bar{Y}^2$

$$\alpha(Y^{(2)} - \bar{Y}^2) = \bar{Y}^2$$

$$\hat{\alpha} = \frac{n\bar{Y}^2}{\sum Y_i^2 - n\bar{Y}^2} \qquad \hat{\beta} = \frac{\sum Y_i^2 - n\bar{Y}^2}{n\bar{Y}}.$$

Now, $\bar{Y}$ and $\frac{1}{n}\sum Y_i^2$ converge in probability, so so do $\hat{\alpha}$ and $\hat{\beta}$. Hence these are consistent estimators.

The Gamma likelihood is $\propto y^{\alpha-1}\exp[-y/\beta] =$

$= \exp\left[(\alpha-1)\log y - \frac{1}{\beta}\cdot y\right]$ w prop. constant dep. on α, β.

So it's an exponential family w $T(y) = \begin{pmatrix}\log y \\ y\end{pmatrix}$.

The joint likelihood has as minimal sufficient statistics:

$\Sigma \log y_i$, Σy_i or equivalently Πy_i, Σy_i.

Using these, we could find better estimators than $\hat{\alpha}, \hat{\beta}$ — but it would take considerably more effort.

Example $Y_1, \dots, Y_n$ iid with density $\begin{cases} (\vartheta+1) y^{\vartheta} \\ \text{for } 0<y<1, \\ \vartheta > -1. \end{cases}$

Find estimator for ϑ.

$$\mathbb{E} Y = \int y f(y)\,dy = \int_0^1 (\vartheta+1) y^{\vartheta+1}\,dy = \frac{\vartheta+1}{\vartheta+2}$$

So if $\mu = \dfrac{\vartheta+1}{\vartheta+2}$ then $\mu(\vartheta+2) = \vartheta+1$ so $\vartheta(\mu-1) = 1-2\mu$

So $\hat{\vartheta} = \dfrac{1-2\bar{x}}{\bar{x}-1}$ is a method of moments estimator.

It is not a function of $\Sigma \ln Y_i$ or of ΠY_i, so it can be improved using Rao-Blackwell.

Example $Y_1, \ldots, Y_n \sim \text{Poisson}(\lambda)$ iid.

The Poisson distribution has mean λ, so $\bar{X}$ is a method of moments estimator.

Example $X_1, X_2, \ldots \sim \text{Bernoulli}(p)$.

Let Y be the index of the first success.
This makes $Y \sim \text{Geometric}(p)$, with
$P(y) = (1-p)^{y-1} p$.

The geometric distribution has mean $\frac{1}{p}$.
Hence $\frac{1}{\bar{Y}}$ is a method of moments estimator of p.

Example $Y_1, \ldots, Y_n \sim \text{Uniform}(0, 3\theta)$ iid.
Then the mean is 1.5θ. So $\frac{\bar{Y}}{1.5}$ is a method of moments estimator.

Maximum Likelihood

Method of Moments is easy, but tends to not produce MVUEs.

A different approach, more likely to produce MVUEs, is to pick

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(x|\theta)$$

the maximizer of the likelihood function.

Example $Y_1, \dots, Y_n \sim \text{Bernoulli}(p)$ iid.

We know that

$$\mathcal{L}(y|p) = p^y (1-p)^{n-y} \quad \text{where } y = \Sigma y_i$$

$$\frac{d\mathcal{L}}{dp} = y p^{y-1} (1-p)^{n-y} + p^y (1-p)^{n-y-1} \cdot (n-y) \cdot (-1)$$

So $\frac{d\mathcal{L}}{dp} = 0$ when

$$y p^{y-1} (1-p)^{n-y} = (n-y) p^y (1-p)^{n-y-1}$$

$$y(1-p) = (n-y)p$$

$$y - yp = (n-y)p$$

$$y = np$$

Which yields the estimator $\hat{p} = y/n$.

Example $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ iid

$$\mathcal{L}(y|\mu,\sigma^2) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i-\mu)^2}{2\sigma^2}\right] =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}} \exp\left[\frac{-1}{2\sigma^2} \Sigma (y_i-\mu)^2\right].$$

Since logarithm is monotonous, $\mathcal{L}$ is maximized when $\ell = \log \mathcal{L}$ is maximized. So we can look at

$$\ell = \log \mathcal{L} = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \Sigma (y_i-\mu)^2$$

Taking derivatives we get

$$\frac{d\ell}{d\mu} = \frac{1}{\sigma^2} \Sigma (y_i-\mu) \qquad = 0 \text{ when } \Sigma y_i = n\hat{\mu}$$
$$\text{or } \hat{\mu} = \bar{y}.$$

$$\frac{d\ell}{d\sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \Sigma (y_i-\mu)^2$$

$$= 0 \text{ when } \frac{1}{2\hat{\sigma}^4} \Sigma (y_i-\hat{\mu})^2 = \frac{n}{2\hat{\sigma}^2}$$

$$\Sigma (y_i-\hat{\mu})^2 = n\hat{\sigma}^2$$

$$\text{or} \qquad \hat{\sigma}^2 = \frac{\Sigma (y_i-\hat{\mu})^2}{n}.$$

Notice: $\hat{\sigma}^2$ is biased! Easy to correct, but still we should be aware that MLE can produce biased estimators.

Example (cont...)

Same setup would for the method of moments yield:

$$\bar{X} = \mu \quad \text{and} \quad \frac{\sum X_i^2}{n} = \sigma^2 + \mu^2$$

So $\bar{x}$ and $\dfrac{\sum X_i^2 - n\bar{x}^2}{n} = \dfrac{\sum (X_i - \bar{X})^2}{n}$ would be the resulting estimators.

//

Using the factorization criterion for a sufficient statistic, we can note that for varying ϑ, the term $h(x)$ is constant. Hence maximizing $\mathcal{L}$ is the same as maximizing $g(\vartheta, u)$

Hence the MLE will usually be a function of the minimal sufficient statistic, and therefore will tend to produce MVUE estimators (after bias correction).

//

MLEs are invariant: if $\hat{\vartheta}$ is MLE for ϑ, then for function t, $t(\hat{\vartheta})$ is the MLE for $t(\vartheta)$.

Example $Y_1, \dots, Y_n \sim \text{Uniform}(0, \vartheta)$.

$$L(Y_i|\vartheta) = \frac{1}{\vartheta^n} \mathbb{1}_{[-\infty, \vartheta]}(\max) \cdot \text{stuff}$$

$\frac{1}{\vartheta^n}$ is decreasing, so maximized at the smallest "valid" value. So the MLE is $\hat{\vartheta} = \max$.

Example $Y_1, \dots, Y_n \sim \text{Poisson}(\lambda)$.

$$L = \prod \frac{\lambda^{Y_i} e^{-\lambda}}{Y_i!} = \lambda^{\sum Y_i} e^{-n\lambda} \cdot h(y) \text{ maximized when } \lambda^{\sum Y_i} e^{-n\lambda}.$$

$$\ell = \log \lambda \cdot \sum Y_i - n\lambda$$

$$\frac{d\ell}{d\lambda} = \frac{\sum Y_i}{\lambda} - n \qquad = 0 \text{ when } \frac{\sum Y_i}{\lambda} = n$$

$$\frac{\sum Y_i}{n} = \bar{Y} = \hat{\lambda}$$

Example $Y \sim \text{Geometric}(p)$.

$$L = p(1-p)^{Y-1}$$

$$\frac{dL}{dp} = (1-p)^{Y-1} + p \cdot (Y-1)(1-p)^{Y-2} \cdot (-1)$$

$$= 0 \text{ when } (1-p)^{Y-1} = p(Y-1)(1-p)^{Y-2}$$

$$1-p = pY - p$$

$$1 = pY$$

So $\hat{p} = 1/Y$ is MLE.