**Introductory Example**

**Slightly Bent Coin.** If the coin is tossed \( n \) times and \( X \) is the number of heads, then

\[
P[X = x] = \binom{n}{x} \theta^x (1 - \theta)^{n-x}
\]

where \( \theta \) is the probability of heads, and

\[
\hat{\theta}_{\text{MLE}} = \hat{\theta}_{\text{MOM}} = \frac{X}{n}
\]

is unbiased and efficient. Works well with large \( n \).

**Small \( n \).** If \( n = 3 \), then the possible values of \( \hat{\theta} \) are 0, 1/3, 2/3, and 1. This is suspect if the coin is only slightly bent.

**Prior Information (or Opinion):** \( \theta \) is thought to be close to 1/2.

---

**Bayesian Inference Estimation**

- Statistical Model with unknowns \( \theta \), say \( f(x; \theta) \).
- Prior distribution for \( \theta \), say \( g(\theta) \), often subjective.
- Compute posterior (conditional) distribution, \( g(\theta|x) \).
- Bayes estimate is the mean of the posterior distribution,

\[
\int \theta g(\theta|x) d\theta.
\]

---

**Beta Distributions**

**The Density.** The beta distribution with parameters \( a > 0 \) and \( b > 0 \) has density

\[
\Gamma(a+b) \frac{\theta^{a-1} (1-\theta)^{b-1}}{\Gamma(a) \Gamma(b)}
\]

for \( 0 < \theta < 1 \), where \( \Gamma \) is the Gamma Function

\[
\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.
\]

**The mean and Variance** of this distribution are

\[
\mu = \frac{a}{a+b} \quad \text{and} \quad \sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}
\]

**Note:** Uniform if \( a = b = 1 \).
Bayesian Inference

**Statistical Model:** $X$ has density or mass function $f(x; \theta)$.

**Prior Distribution:** $\theta$ has a prior density $g$, typically subjective.

**Other Densities:** the joint density of $\theta$ and $X$ is

$$g(\theta) \times f(x; \theta) = L(\theta|x)g(\theta).$$

The marginal density of $X$ is

$$c(x) = \int L(\theta|x)g(\theta)d\theta;$$

and the conditional (posterior) density of $\theta$ given $x$ is

$$\frac{L(\theta|x)g(\theta)}{c(x)} \propto L(\theta|x)g(\theta). \quad (*)$$

**Note:** $(*)$ is Bayes Theorem.

---

**Example: Binomial**

Suppose $X \sim \text{Binomial}(n, \theta)$ and $\theta \sim \beta(a, b)$. Then

$$L(\theta|x) = f(x; \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x},$$

$$g(\theta|x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1},$$

$$g(\theta|x) = L(\theta|x)g(\theta)$$

$$= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a+x-1}(1-\theta)^{b+n-x-1}.$$

---

**Mathematical Interlude**

If $g_1$ and $g_2$ are two densities for which

$$g_1(\theta) = cg_2(\theta)$$

for all $\theta$, then $c = 1$ and $g_1 = g_2$, since

$$1 = \int g_2(\theta)d\theta = c \int g_1(\theta)d\theta = c.$$

**Note:** So, if $g_1 \propto g_2$, then $g_1 = g_2$.

That is,

$$g(\theta|x) \propto \theta^{a'-1}(1-\theta)^{b'-1}$$

where $a' = a + x$ and $b' = b + n - x$. So

$$\theta|x \sim \beta(a' = a + x, b' = b + n - x),$$

and the Bayes estimate is

$$\frac{a'}{a' + b'} = \frac{a + x}{a + b + n}.$$

**Notes**

1. Revised opinion: $a$ and $b$ are changed to $a'$ and $b'$.

2. Bayes estimate combines prior information ($a$ and $b$) with the data ($x$).
Numerical Examples 1. If $a = b = 2$, $n = 1000$, and $x = 505$, then the Bayes estimate and MLE are both .505 (to three decimals).

2. If $a = b = 9$, $n = 3$, and $x = 0$, then the Bayes estimate is .429 and the MLE is 0.

3. If $a = b = 9$, $n = 20$, and $x = 8$, then the Bayes estimate is .448 and the MLE is .400.

Notes 1. If $a + b << n$, then the Bayes estimate is approximately the MLE.

2. If $n << a + b$, then the Bayes estimate is approximately the prior mean.

Conjugate Priors

Beta-Binomial: There was special structure in the example: the prior and posterior distributions were both beta distributions, but with different parameters.

Gamma Poisson. If $X_1, \cdots, X_n \sim \text{indPoisson}(\theta)$ and $\theta$ has a prior gamma distributions, with density

$$b^a \theta^{a-1} e^{-b \theta} / \Gamma(a),$$

then the posterior is again a gamma distribution with new parameters

$$a' = a + x_1 + \cdots + x_n \quad \text{and} \quad b' = b + n.$$

Example: Setting a Fair Auto Insurance Rate

Cover Story: An insurance company insures many drivers and has data on past performance. For a given driver, let

$$X = \# \text{ accidents last year},$$
$$Y = \# \text{ accidents next year},$$

and suppose that $X, Y \sim \text{ind Poisson}(\theta)$, where $\theta$ depends on the driver. Here $X$ is observed, but not $Y$, and interest centers on $E(Y) = \theta$.

Question: What is $E(\theta|x)$. 

Normal Normal: If $X_1, \cdots, X_n \sim \text{ind Normal}(\mu, \sigma^2)$, where $\sigma^2$ is known, and $\mu$ has a prior $\text{Normal}(\mu, \tau^2)$ distribution, then the posterior distribution is normal with

$$\mu' = \frac{\tau^{-2} \mu + n \sigma^{-2} \bar{x}}{\tau^{-2} + n \sigma^{-2}}$$

and

$$\sigma'^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + n \tau^2}.$$

Note: See the text for the derivation.
**The Prior:** Suppose that the mean and standard deviation of the number of accidents among all drivers last year were

\[ \mu = .025 \quad \text{and} \quad \sigma = .05. \]

Recall the the mean and standard deviation of the gamma distribution are

\[ \mu = \frac{a}{b} \quad \text{and} \quad \sigma = \frac{\sqrt{a}}{b}. \]

Thus, if \( a = .25 \) and \( b = 10 \), then \( \mu = .025 \) and \( \sigma = .05 \).

**Compromise:** Suppose that \( \theta \sim \Gamma(a = .25, b = 10) \).

---

**The Conditional Expection.** For a single driver, the likelihood function is

\[ L(\theta|x) = \frac{1}{x!} \theta^x e^{-\theta}. \]

If the prior is

\[ g(\theta) = \frac{b^a \theta^{a-1}}{\Gamma(a)} e^{-b\theta} \]

then

\[ g(\theta|x) \propto L(\theta|x)g(\theta) \propto \theta^{a+x-1} e^{-(b+1)\theta}. \]

So the posterior distribution of \( \theta \) is a gamma distribution with \( a' = a + x \) and \( b' = b + 1 \); and the posterior mean is

\[ E(\theta|x) = \frac{a'}{b'} = \frac{a + x}{b + 1}. \]

---

**Numerically:** For \( a = .25 \) and \( b = 10 \), this is

\[ E(\theta|x) = \frac{.25 + x}{11}. \]

**Some Values**

| \( x \) | \( E(\theta|x) \) |
|-------|----------------|
| 0     | .0227          |
| 1     | .1136          |
| 2     | .2045          |
| 3     | .2955          |

**Note:** Small effect for \( x = 0 \); large for \( x > 0 \).

---

**Bias, Variance, and MSE**

**An Example: The Normal Case**

**Priors and Posteriors:** If \( X \sim \text{Normal}(\theta, \sigma^2) \) and \( \theta \sim \text{Normal}(\mu, \tau^2) \), then

\[ \theta|x \sim \text{Normal}(\mu', \tau'^2), \]

where

\[ \mu' = \frac{\sigma^2 \mu + \tau^2 x}{\sigma^2 + \tau^2} \quad \text{and} \quad \tau'^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}. \]

So, the Bayes estimate of \( \theta \) is \( \mu' \). Here

\[ E(\mu'|\theta) = \frac{\sigma^2 \mu + \tau^2 \theta}{\sigma^2 + \tau^2}. \]
Recall:
\[
\mu' = \frac{\sigma^2 \mu + \tau^2 x}{\sigma^2 + \tau^2} \quad \text{and} \quad E(\mu'|\theta) = \frac{\sigma^2 \mu + \tau^2 \theta}{\sigma^2 + \tau^2}.
\]
So,
\[
\text{bias} = \frac{\sigma^2 \mu + \tau^2 \theta}{\sigma^2 + \tau^2} - \theta = \frac{\sigma^2 (\mu - \theta)}{\sigma^2 + \tau^2}
\]
and
\[
\text{Var}(\mu') = \left(\frac{\tau^2}{\sigma^2 + \tau^2}\right)^2 \sigma^2.
\]
So,
\[
\text{MSE} = \left(\frac{\tau^2}{\sigma^2 + \tau^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 (\theta - \mu)^2.
\]

Comparisons
The MLE $X$ is unbiased and its MSE is $\sigma^2 + 0 = \sigma^2$. So, the MSE of the Bayes estimator is less than the MSE of the MLE is
\[
\left(\frac{\tau^2}{\sigma^2 + \tau^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right)^2 (\theta - \mu)^2 < \sigma^2
\]
or equivalently
\[
|\theta - \mu| \leq \sqrt{\sigma^2 + 2\tau^2}.
\]
So, if $\theta$ really does fall under the prior density, then the Bayes estimator is better.

Figure 2: Beta density with $a = b = 9$