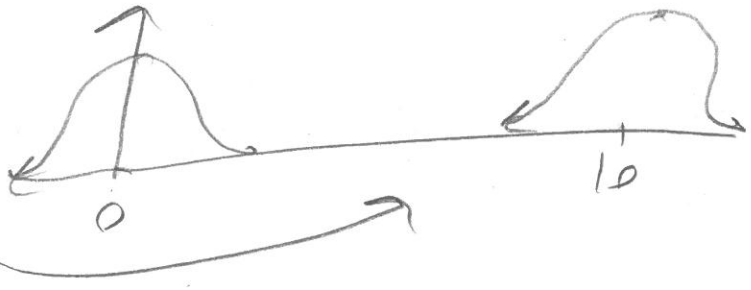
Lecture 7 2/29/16

If

$$X \sim \begin{cases} N(0,1^2) & \text{w.p. } \frac{1}{2} \\ N(10,1^2) & \text{w.p. } \frac{1}{2} \end{cases}$$



What is PDF of X?

$= \frac{1}{2} N(0,1) + \frac{1}{2} N(10,1^2)$ ← we will return to this later

$$p(x) = \sum_{\theta \in \{H\}_0} p(x;\theta)\, p(\theta) = \sum_{\theta \in \{0,10\}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \left(\frac{1}{2}\right)$$

$$= \frac{1}{2\sqrt{2\pi}} \left(e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}(x-10)^2} \right)$$

Okay... what about

$$X \sim \begin{cases} \text{Bin}(n, 0.25) & \text{w.p. } 0.1 \\ \text{Bin}(n, 0.4) & \text{w.p. } 0.9 \end{cases}$$

"marginal distribution" or "compound distr."

$$p(x) = \sum_{\theta \in \{25, .4\}} \binom{n}{x} \theta^x (1-\theta)^{n-x} p(\theta)$$

$$= \binom{n}{x} \left(.25^x .75^{n-x} \cdot 0.1 + .4^x .6^{n-x} \cdot 0.9 \right)$$

What if

$X \sim Bin(n, \theta)$ but $\theta \sim U(0,1)$

(RRR GGG) (RGG GGG) (RRR RRG) . . .

Bags of marbles each with diff prop of Red's

& bags form a $U(0,1)$

$$p(x) = \int_{\theta \in (0,1)} f(x;\theta)\, p(\theta)\, d\theta = \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} (1)\, d\theta = \binom{n}{x} B(x+1, n-x+1)$$

this is a dist'r!

$$\sum_{x=0}^{n} \binom{n}{x} B(x+1, n-x+1) = 1$$

→ will see who it's called soon

Wait what does this look like?

derivation in bags

What about $X \sim Bin(n, \theta)$ but $\theta \sim Beta(\alpha, \beta)$?

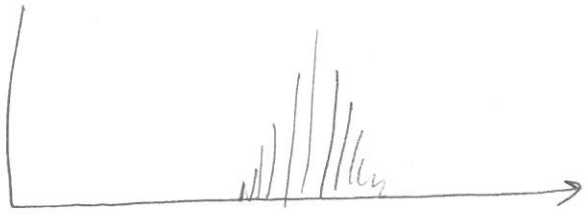
$$p(x) = \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}\, d\theta = \frac{\binom{n}{x}}{B(\alpha,\beta)} \int_0^1 \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}\, d\theta$$

$$= \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)}$$

"BetaBin(n, α, β)"

Known as the beta-binomial distribution. $E(X)$? $n\frac{\alpha}{\alpha+\beta}$ $Var(X) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$X \sim Bin(n, \theta)$ vs $X \sim BetaBin(n, \alpha, \beta)$ assume n the same.



Bin has one param

$\Rightarrow E(X) = n\theta$
$Var(X) = n\theta(1-\theta)$] pick mean $\Rightarrow$ var fix

BetaBin has two params $\Rightarrow$ you can pick mean & variance, you can get "burstiness" or "overdispersion" as above or very low variance

If $\alpha, \beta \to \infty$ it can rise

$\lim E(X) = n\left(\frac{1}{2}\right)$

$\lim Var(X) = n \lim_{\alpha\to\infty} \frac{\alpha^2(2\alpha+n)}{(2\alpha)^2(2\alpha+1)} = n\frac{1}{4}$

$\Rightarrow \lim_{\alpha,\beta\to\infty} BetaBin(n,\alpha,\beta) = Bin(n, \frac{1}{2})$

This makes sense because: Likewise $\lim_{\alpha,\beta\to\infty} \frac{\alpha}{\beta} = c$ you can get $Bin(n,\theta)$ that just is $Bin(n\theta)$ when θ is known from dataset dist.

The overdispersion gives you flexibility in the case where you're unsure about θ!

e.g. Wikipedia. Families with 13 children. Gender of first 12

children	# of males	0	1	2	3	4	5	6	7	8	9	10	11	12	
	X	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	36,100 families
=6115 families	$\hat{X}$ BetaBin$(n,39,32)$	2	23	95	311	656	1036	1258	1187	854	462	170	44	5	
$\Sigma x =$	$\hat{X} \sim$ Bin$(n, .519)$	1	12	72	259	626	1085	1367	1366	1054	610	232	56	2	$\Rightarrow$ 519

Beta Bin fits better! Why is human gender ratio beta-binomial families??

Recall: $X_1, \dots X_n \overset{\text{exch}}{\sim} \text{Bern}(\theta)$, $\theta \sim \text{Beta}(\alpha, \beta)$

$\Rightarrow \theta | X \sim \text{Beta}(\alpha + x, \beta + n - x)$ & $X^* | X \sim \text{BetaBin}\left(n, \frac{x+\alpha}{n+\alpha+\beta}\right)$

Jeffreys prior $X \sim \text{BetaBin}(n, \alpha, \beta) \Rightarrow$ prior predictive distribution for dataset of size n

---

$X^* | X \sim$???

↑

let X^* def length m.

We are predicting the outcome not of only the very next trial ... but of the next m trials given n data pts and a Beta prior on θ.

$$p(x^* | x) = \int_{\theta \in (0,1)} \overbrace{p(x^* | \theta)}^{\text{Bin}(m,\theta)} \overbrace{p(\theta | x)}^{\text{Beta}(\alpha^*, \beta+n-x)} d\theta$$

$$= \int_0^1 \binom{m}{x^*} \theta^{x^*} (1-\theta)^{m-x^*} \frac{1}{B(\alpha+x, \beta+n-\alpha)} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1} \, d\theta$$

$$= \frac{\binom{m}{x^*}}{B(\alpha+x, \beta+n-\alpha)} \int_0^1 \theta^{x^*+x+\alpha-1} (1-\theta)^{m-x^*+n-x+\beta-1} \, d\theta$$

$$= \frac{\binom{m}{x^*}}{B(\alpha+\alpha, \beta+n-\alpha)} B\left(x^*+x+\alpha,\ m-x^*+1-\alpha+\beta\right)$$

$$= \text{BetaBin}\left(m, x+\alpha, n-x+\beta\right)$$

How; prove for $m=1$ this is a Beta

Why should this be?

If θ was precisely known

$$X^*|x = X^*|\theta \sim \text{Bin}(m, \theta)$$

But θ is known with uncertainty and other uncertainty ~ Beta, it is as if on each trial, you grab a $\theta|x$ and use that to run a Bernoulli, then grab another one, etc...

FACT:

Posterior predictive does correctly / fractionally incorporates uncertainty in your best guess of θ!!

Review.... $X_1, \dots, X_k$ each $\text{Bern}(\theta)$, or $\text{Beta}(\alpha, \beta)$

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \propto P(x|\theta)P(\theta)$$

↑ minus the $\frac{1}{P(x)}$ zoom factor.

Note that it's now a function of θ

$$P(\theta|x) \propto \binom{k}{x} \theta^x (1-\theta)^{k-x} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} = \frac{\binom{k}{x}}{B(\alpha,\beta)} \theta^{x+\alpha-1}(1-\theta)^{k-x+\beta-1}$$

$$\propto \theta^{x+\alpha-1}(1-\theta)^{k-x+\beta-1}$$

$P(\theta|x) = P(\theta|x;\alpha,\beta)$

↑↑↑

assuming fixed "hyperparameters"

"above"

Why?

$\forall\theta$ this is a constant

$\forall\theta$ this is a constant. But what about x? $P(\theta|x)$. x is fixed!!

That's what conditional prob means!

Compound distr: marginalizing over the intermediate parameter

Here the result is a prob of hyperparameters

$$P(x;\alpha,\beta) = \int_{\theta\in(0,1)} P(x|\theta)P(\theta;\alpha,\beta)\,d\theta$$

$$P(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{k-x+\beta-1}$$

$$\Rightarrow \frac{P(\theta_1|x)}{P(\theta_2|x)} = \left(\frac{\theta_1}{\theta_2}\right)^{x+\alpha-1}\left(\frac{1-\theta_1}{1-\theta_2}\right)^{k-x+\beta-1}$$

You can get posterior odds ratios!!

But can you get the whole density? Yes...

$$p(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

If I integrate this... what do I get?

$$\int_0^1 d\theta \neq 1 \quad \text{but} < \infty$$

$$\Rightarrow = c$$

$$\int g(x,\theta)\,d\theta = c \Rightarrow f(\theta) = \frac{g(x,\theta)}{c}$$ so it is now a density

Hold on... $X \sim Bin(n,\theta)$

$$p(x|\theta,n) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \frac{n!}{(n-x)!\,x!}\theta^x(1-\theta)^n(1-\theta)^{-x}$$

$$\propto \underbrace{\frac{1}{(n-x)!\,x!}\left(\frac{\theta}{1-\theta}\right)^x}_{\text{Kernel of the PMF/PDF}}$$ this doesn't look like a binomial!!

$$f(x;\theta) = \underbrace{k(\theta)}_{\text{normalizing constant}}\underbrace{g(x,\theta)}_{\text{"Kernel"}}$$

From a kernel, the normalizing constant is fixed - therefore a r.v. can be identified by its kernel!

$\alpha, \beta \in (0, \infty)$

$\theta \sim \text{Beta}(\alpha, \beta) := \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$= \theta^{a}(1-\theta)^{b} \quad a, b \in (-1, \infty)$

Anything of this form is a beta!

$p(\theta|x) \propto \theta^{\overbrace{x+\alpha-1}^{a}} (1-\theta)^{\overbrace{n-x+\beta-1}^{b}}$ → the kernel for a beta distr.

$\Rightarrow \theta|x \sim \text{Beta}(a, b)$

How: you will find the kernels for all the common distr's.

Bayes-Laplace prior

$\theta|x \sim \text{Beta}(\alpha+x, \beta+n-x)$, $\theta \sim U(0,1) = \text{Beta}(1,1)$

"Wilson Estimate"

$\hat{\theta}_{mmse} = \frac{\alpha+x}{\alpha+\beta+n} = \frac{x+1}{n+2}$ } "Law of Succession". Very famous estimator of θ. Highly depend only because Bayesian stat is highly disputed or

$\hat{\theta}_{map} = \frac{x}{n} = \hat{\theta}_{MLE}$

(interesting)

He was working on the Sunrise problem pre found by Hume.

Recall the post. distr.:

$$\theta | x \sim \text{Beta}(\alpha + x, \beta + n - x)$$

$$\hat{\theta}_{MMSE} = E[\theta | x] = \frac{\alpha + x}{\alpha + \beta + n} = \frac{\alpha}{\alpha+\beta+n} + \frac{x}{\alpha+\beta+n} \quad // \quad \frac{\alpha}{\alpha+\beta+n}\left(\frac{\alpha+\beta}{\alpha+\beta}\right) + \frac{x}{\alpha+\beta+n}\left(\frac{n}{n}\right)$$

$E[\theta]$ prior mean $\qquad \hat{\theta}_{MLE} = \bar{x}$

$$= \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta}\right) + \frac{n}{\alpha+\beta+n}\left(\frac{x}{n}\right)$$

Note $\frac{\alpha+\beta}{\alpha+\beta+n} + \frac{n}{\alpha+\beta+n} = 1$

let $\rho := \frac{\alpha+\beta}{\alpha+\beta+n}$ $\Rightarrow 1-\rho = \frac{n}{\alpha+\beta+n}$ and both are prop's which sum to 1.

$$\Rightarrow \hat{\theta}_{MMSE} = \rho E[\theta] + (1-\rho)\hat{\theta}_{MLE}$$

Where ρ is known as the shrinkage factor / proportion.

$\hat{\theta}_{MMSE}$ is known as a "shrinkage estimator"

Since it "shrinks" towards your prior best guess, $E[\theta]$.

Large values of ρ shrink more toward ρ and small values of ρ let the data speak for itself. (If $n \uparrow$, $\rho \downarrow$ $\Rightarrow$ data speaks for itself)

$$\text{Beta}(\alpha+x, \beta+n-x) \propto \theta^{(\alpha-1)+(x)} (1-\theta)^{(\beta-1)+(n-x)} \quad (\text{Hw 3 \#3c})$$

(# prior successes, # successes, # failures)

"Bayes prior"

So $\theta \sim U(0,1) = \text{Beta}(1,1)$ is equivalent to seeing no successes or failures

but $\rho = \frac{2}{2+n} \to 0$ as $n \to \infty$ but is still shrunk toward $E[\theta] = \frac{1}{2}$