 see his text for proofs.

Lemma 6.1: Gamma Function Property

$$\frac{\Gamma(x+1)}{\Gamma(x)} = x$$

Latex version coming soon!

Definition: Beta Density Function

$$\rho(f) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} f^{a-1}(1-f)^{b-1} \qquad 0 \le f \le 1$$

Lemma 6.2: Relating Gamma and Beta

$$\int_0^1 f^a(1-f)^b \, df = \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \qquad a,b \in \mathbb{R}_{>0}$$

Lemma 6.3:

$E(F) = a/N$, $a, b, N = a+b$ parameters of a beta distribution for F

Proof done using Lemmas 6.1, 6.2

$E(F)$ is "expected value", and is defined to be our "estimate of the relative freq."
we further assume our beliefs such that:

$P(X=1 \mid f) = f$ That is, if we know for a fact that the relative frequency with which $X=1$ was f, our belief concerning the occurrence of 1 in the first execution of the experiment would be f.

Note:

Many use (θ) theta for a random variable whose value is a relative frequency. F is used here since capital/lowercase theta are similar and easy to confuse.

Theorem 6.1

Suppose X is a random variable with space $\{1, 2\}$ (values), and F is another random variable such that

$P(X=1 \mid f) = f$, then

$P(X=1) = E(F)$

Proof:

$$P(X=1) \underset{\text{Law of total probability}}{=} \int_0^1 P(X=1 \mid f)\rho(f)\,df = \int_0^1 f\rho(f)\,df = E(F) \;\square$$

Corollary 6.1: if F is beta with parameters a, b, $N = a+b$,

$$P(X=1) = a/(a+b) = a/N$$

(by Theorem 6.1 and Lemma 6.3)

Definition 6.2:

D is called a sample of size M with parameter F if:

- $D = \{X^{(1)}, X^{(2)}, \ldots, X^{(M)}\}$ s.t. each $X^{(h)}$ has the same space.
- There is a random variable F with density function ρ s.t. all $X^{(h)}$'s are mutually independent conditional on F, and $\forall f \in F$, all $X^{(h)}$'s have the same probability distribution conditional on f.

Definition 6.3:

Given a sample of size M such that:

- each $X^{(h)}$ has space $\{1, 2\}$ (same 2 possible values)
- $\mathcal{F} = \{F\}$ (singleton set), F has space $[0,1]$ (probability interval) and for $1 \le h \le M$, $P(X^{(h)} = 1 \mid f) = f$. $(*)$

Then D is called a binomial sample of size M with parameter F

Lemma 6.4:

Given F with beta distribution w/ a, b, $N = a+b$, $s, t \in \mathbb{N}$, $M = s+t$,

$$E(F^s[1-F]^t) = \frac{\Gamma(N)}{\Gamma(N+M)} \frac{\Gamma(a+s)\Gamma(b+t)}{\Gamma(a)\Gamma(b)}$$

Proof: Use fact $E(F) = \int f \rho(f) df$, Lemma 6.2, definition of beta distr.

$$E(F^s[1-F]^t) = \int_0^1 f^s(1-f)^t \rho(f) df \ \ldots \text{ etc.}$$

Lemma 6.5:

Given F with beta with $a, b, N = a+b$, $s, t \in \mathbb{N}$, $M = s+t$, then

$$\frac{f^s(1-f)^t \rho(f)}{E(F^s[1-F]^t)} = \text{beta}(f; a+s, b+t)$$

Proof: use Lemma 6.4 ~~and~~ definition of beta distribution

Theorem 6.2

Given:

- D, a binomial sample of size M with parameter F
- a set of values $d = \{x^{(1)}, x^{(2)}, \ldots, x^{(M)}\}$ of the variables in D, called our "data set" or just "data"
- s, t, the number of variables in $d = 1, 2$, respectively.

Then:

$$P(d) = E\left(F^s[1-F)^t\right)$$

Proof (continuous case)

$$P(d) = \int_0^1 P(d|f)\rho(f)\,df$$

$$= \underbrace{\int_0^1 \prod_{h=1}^{M} P(x^{(h)}|f)\rho(f)\,df}_{\text{variables in } d \text{ are C.I. given } F} = \underbrace{\int_0^1 f^s(1-f)^t \rho(f)\,df}_{\text{Apply } (*) \text{ from Def 6.3 repeatedly}} = E\left(F^s[1-F]^t\right)$$

Corollary 6.2

If F has a beta distribution with $a, b, N=a+b$, $s,t \in \mathbb{N}$, $s+t=M$

$$P(d) = \frac{\Gamma(N)}{\Gamma(N+M)} \frac{\Gamma(a+s)\Gamma(b+t)}{\Gamma(a)\Gamma(b)},$$ Proof by Thm. 6.2, Lemma 6.4.

Theorem 6.3

Given conditions for Thm. 6.2.

$$\rho(f|d) = \frac{f^s(1-f)^t\rho(f)}{E(F^s[1-F)^t)}$$, where $\rho(f|d)$ is the density function of F conditional on $D=d$.

Proof:

$$\rho(f|d) = \underbrace{P(d|f)\rho(f)/P(d)}_{\text{Bayes' theorem}} = \underbrace{\frac{f^s(1-f)^t\rho(f)}{E(F^s[1-F]^t)}}_{\text{Equality } (*) \text{ when variables in } D \text{ are indep. conditional on } F \text{; Thm 6.2}}$$

Corollary 6.3

Given conditions for 6.2, and F is beta w/ $a, b, N=a+b$, $s,t \in \mathbb{N}$, that is, $\rho(f) = beta(f; a, b)$, then $\rho(f|d) = beta(f; a+s, b+t)$.

Proof by Thm 6.3, Lemma 6.5

Theorem 6.4

Suppose conditions for Thm 6.2 hold, and we create a binomial sample of size $M+1$ by adding another variable $X^{(M+1)}$. Then, if D is the binomial sample of size M, the "updated distribution relative to the sample and data d" is

$$P(X^{(M+1)} = 1 \mid d) = E(F \mid d)$$

proof: use equality (*) after seeing that $X^{(M+1)}$ is indep of vars in D cond on F.

Corollary 6.4

if F is beta and conditions of 6.4 hold,

$$P(X^{(M+1)} = 1 \mid d) = \frac{a+s}{N+M}$$

Proof: Thm 6.4, Corollary 6.3, Lemma 6.3

LATEX VERSION COMING SOON!