Simulations: method of composition

- **Model $\pi(X|\theta)$ & uncertain θ:**
  - Usually easy to simulate distribution of X based on **given parameters** θ and the model $\pi(X|\theta)$. And likewise for any transformation g(X).
    - Assuming you can generate samples from $\pi(X|\theta)$.
    - This would not take into account the uncertainty in parameters θ.
    - Assume $\pi(\theta)$ is prior.
    - **Method of composition: simulate a compound distribution: $\pi(X) = \int \pi(X|\theta)\, \pi(\theta)\, d\theta$** which is same as $\int \pi(X,\theta)\, d\theta$ giving the marginal distribution $\pi(X)$.
    - Step 1: simulate θ from $\pi(\theta)$.
    - Step 2: simulate X from $\pi(X|\theta)$ & Go to Step1.

Method of composition: example

- **Model Bin($X|\theta$,N) & uncertain θ:**
  - Step 1: Simulate from prior distribution $\pi(\theta)$ = U(0,1) (for example)
  - Step 2: Simulate X from Bin($X|\theta$).
  - Repeat 1&2, until large sample of X.
  - Plot histogram of X.
  - The resulting distribution shows predictive distribution of X, but not based on any single parameter estimate θ.
  - It is based on uncertain parameter, described as $\pi(\theta)$.
  - Next: replace $\pi(\theta)$ by $\pi(\theta \mid \text{data})$ → data based predictions, accounting for parameter uncertainty!

Posterior predictive distributions

- **Bayesian predictions from data:**
  - Aim: to compute **posterior predictive distribution $P(X_{new} \mid X_{obs})$**
  - This gives prediction based on the past data, not based on assumed parameter estimates.

Predictive distributions

- Consider series of observations: $X_1,\ldots,X_n$ and a model $\pi(X_i \mid \theta)$ so that X_i are conditionally independent, given θ.
  Posterior predictive distribution of X_{n+1} :

$$\pi(X_{n+1} \mid X_1,\cdots,X_n) = \int \pi(X_{n+1},\theta \mid X_1,\cdots,X_n)d\theta$$

$$= \int \pi(X_{n+1} \mid \theta, X_1,\cdots,X_n)\pi(\theta \mid X_1,\cdots,X_n)d\theta$$

$$= \int \underbrace{\pi(X_{n+1} \mid \theta)}_{\text{Our model}}\underbrace{\pi(\theta \mid X_1,\cdots,X_n)}_{\text{Posterior of } \theta}d\theta$$

Predictive distributions

- **Likewise:**
  Prior predictive distribution of X_{n+1} :

$$\pi(X_{n+1}) = \int \pi(X_{n+1}, \theta) d\theta = \int \underbrace{\pi(X_{n+1} \mid \theta)}_{\text{Our model}} \underbrace{\pi(\theta)}_{\text{Prior of } \theta} d\theta$$

- *"With the predictive approach parameters diminish in importance, especially those that have no physical meaning. From the Bayesian viewpoint, such parameters can be regarded as just place holders for a particular kind of uncertainty on your way to making good predictions".* (Draper 1997, Lindley 1972).

Predictive distributions

- Note also, directly from Bayes, we get:

$$\pi(X) = \frac{\pi(X \mid \theta)\pi(\theta)}{\pi(\theta \mid X)}$$

- by inserting *prior*, *posterior*, *model of X*, we find prior predictive density of X.
- Similarly,

$$\pi(X_{n+1} \mid X_1, \cdots, X_n) = \frac{\pi(X_{n+1} \mid \theta, X_1, \cdots, X_n)\pi(\theta \mid X_1, \cdots, X_n)}{\pi(\theta \mid X_1, \cdots, X_{n+1})}$$

Predictive distributions

- **Try exact solving with binomial model.**
  - **Assume we have a posterior which is beta(α,β).**
  - **'Old data' is then included in α,β.**

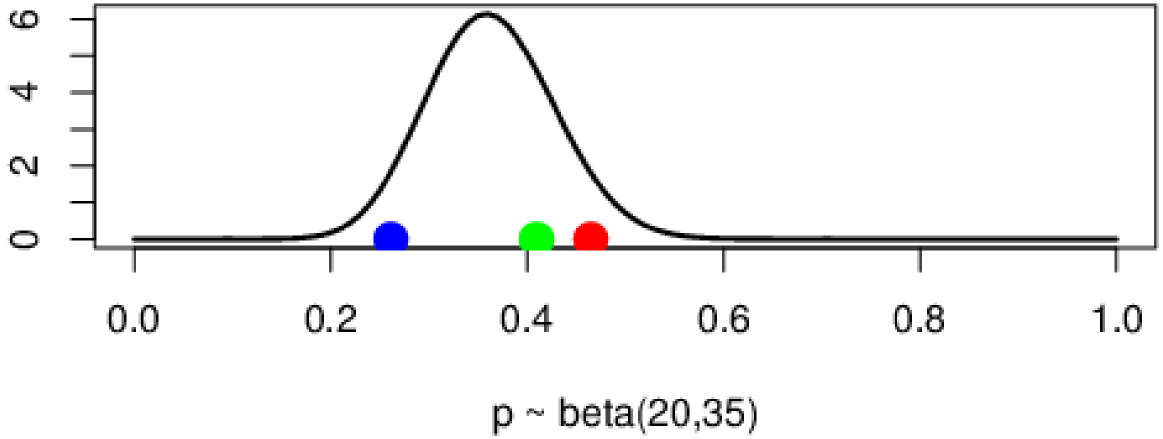
$$\pi(X \mid \alpha,\beta) = \int \pi(X,\theta \mid \alpha,\beta)d\theta = \int \underbrace{\pi(X \mid \theta)}_{\text{Binomial}(N,\theta)}\underbrace{\pi(\theta \mid \alpha,\beta)}_{\text{Beta}(\alpha,\beta)}d\theta$$

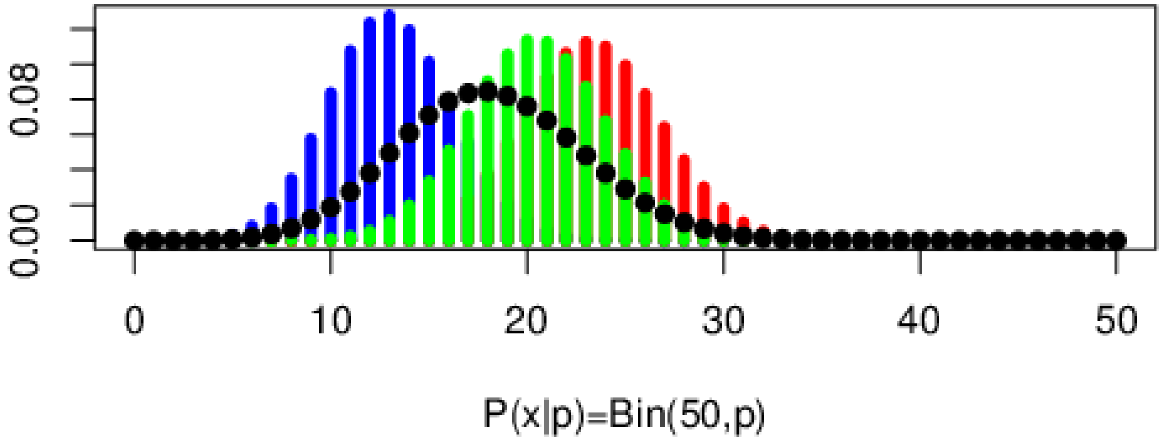
  - **This can be solved as:**

$$\pi(X \mid \alpha,\beta) = \binom{N}{X}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)}$$

A = X+α, B = N-X+β

A BETA-BINOMIAL distribution.


0
2
4
6
0.0
0.2
0.4
0.6
0.8
1.0
p ~ beta(20,35)


0.00
0.08
0
10
20
30
40
50
P(x|p)=Bin(50,p)

Predictive distributions

- **To solve predictive means, variances:**
  - **Use $E(X) = E(E(X|\theta))$**
  - **Use $V(X) = E(V(X|\theta)) + V(E(X|\theta))$**

  - **For example, with Beta & Binomial:**
  - **$E(X) = E(n\theta) = n\,\alpha/(\alpha+\beta)$**
  - **$V(X) = E(n\theta\,(1-\theta)) + V(n\theta) =$**

    $n\alpha/(\alpha+\beta) - nE(\theta^2) + n^2\alpha\beta\,/[(\alpha+\beta)^2(\alpha+\beta+1)]$
    where $E(\theta^2)=V(\theta)+E^2(\theta)$, which leads to

    $= n\alpha\beta\,(\alpha+\beta+n)\,/[(\alpha+\beta)^2(\alpha+\beta+1)]$
  - **By including parameter uncertainty $\pi(\theta)$ to a model $\pi(X|\theta)$ we get models $\pi(X) = \int\pi(X|\theta)\pi(\theta)d\theta$, suitable for e.g. overdispersed data.**