Statistics 360/601 – Modern Bayesian Theory

Alexander Volfovsky

Lecture 8 - Sept 25, 2018

Monte Carlo

Monte Carlo approximation

Want to compute

$$E[\theta|y] = \int \theta p(\theta|y) d\theta$$

without worrying about actual integration...

- Let θ be a parameter of interest
- Let $y_1, \ldots, y_n$ be numerical values of a sample from a distribution $p(y_1, \ldots, y_n|\theta)$
- Let $\theta^1, \ldots, \theta^S \sim p(\theta|y_1, \ldots, y_n)$ be iid samples from the posterior.
- We estimate our posterior quantity of interest as $1/S \sum g(\theta^i)$

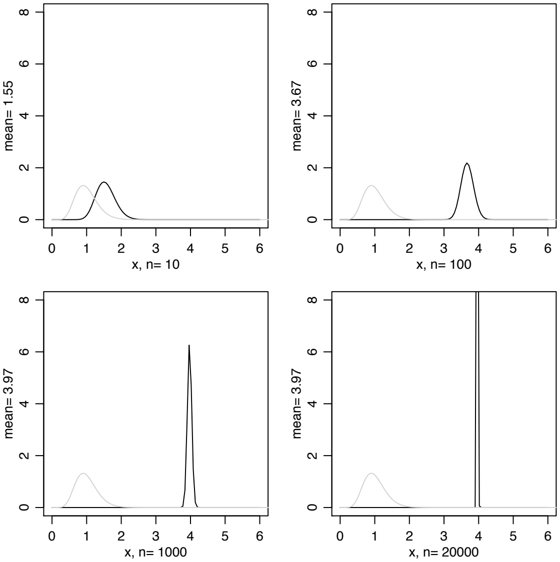
Consistent parameters??

- Let the data come from $Y \sim$Geometric(p).
- Recall that a geometric variable has pdf

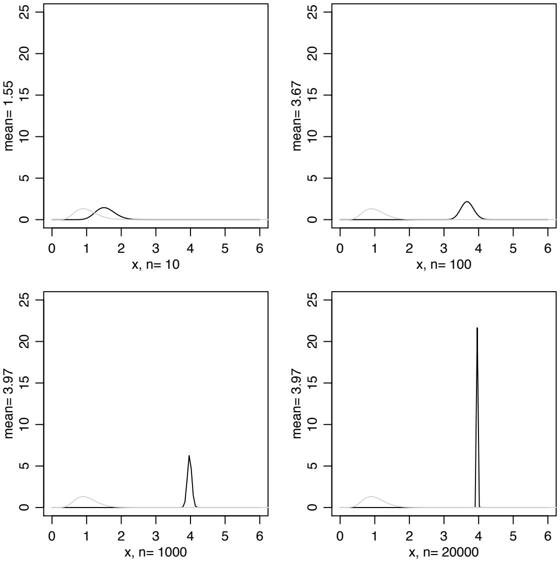
$$\Pr(Y = k) = (1 - p)^k p$$

  and it captures a success on the $k + 1$ trial after k failures.
- The mean of a geometric is $(1 - p)/p$.
- Let the model we think the data comes from be *Poisson*(θ).
- Let the prior be a Gamma(α, β).
- We know the posterior is Gamma($\alpha + \sum y_i, \beta + n$).

Consistent parameters??



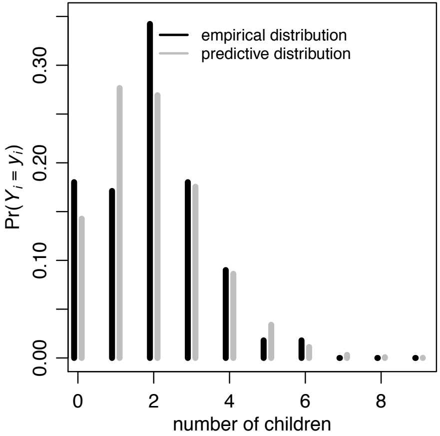
Consistent parameters??



Posterior Predictive Checks

Posterior predictive checking

Lets look at the data (from the book).

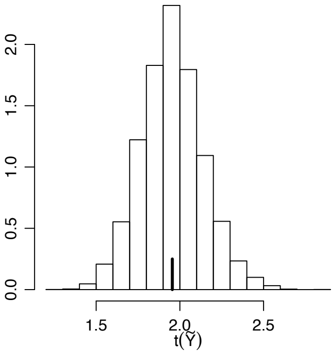


Data: twice as many women with 2 children as with 1.

Posterior predictive: fewer women with 2 children than with 1.

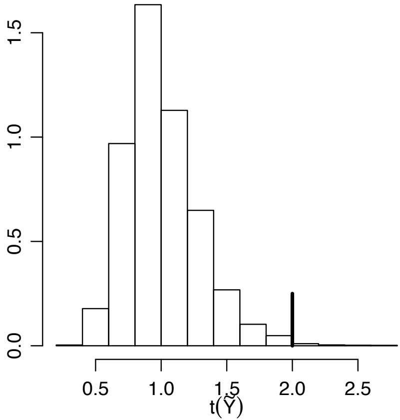
Posterior predictive checking

```
> t.mc <- t2.mc <-NULL
> for(s in 1:10000) {
>  theta1<-rgamma(1,a+sum(y1), b+length(y1))
>  y1.mc<-rpois(length(y1),theta1)
>  t.mc <- c(t.mc,mean(y1.mc))
>  t2.mc<-c(t2.mc,sum(y1.mc==2)/sum(y1.mc==1))
> }
```



Where $t(y)$ is the mean!

Posterior predictive checking



Where $t(y)$ is the ratio of 2's to 1's in a dataset.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.
- n bins of counts based on positions of detectors — 6 million counts.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.
- n bins of counts based on positions of detectors — 6 million counts.
- Bin count y_i are modeled as independent Poisson(θ_i)

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.
- n bins of counts based on positions of detectors — 6 million counts.
- Bin count y_i are modeled as independent Poisson(θ_i)
- $\Theta = Ag + r$ where g is the unknown image, A is a known linear operator and r are known corrections.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.
- n bins of counts based on positions of detectors — 6 million counts.
- Bin count y_i are modeled as independent Poisson(θ_i)
- $\Theta = Ag + r$ where g is the unknown image, A is a known linear operator and r are known corrections.
- A, g and r are non-negative.

Posterior predictive checking

- Lets look at the data (Gelman, Meng and Stern).
- Gelman (1990,92) describe positron emission tomography experiment.
- Goal: Estimate the density of a radioactive isotope in a cross-section of the brain.
- 2-d image is estimate from gamma-ray counts in a ring of detectors around the head.
- n bins of counts based on positions of detectors — 6 million counts.
- Bin count y_i are modeled as independent Poisson(θ_i)
- $\Theta = Ag + r$ where g is the unknown image, A is a known linear operator and r are known corrections.
- A, g and r are non-negative.
- This is an easy problem without the constraint.

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y; \hat{\theta}) \approx 30,000$.

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y;\hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y;\hat{\theta}) \approx 30,000$.
- Reject the model!

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y;\hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y;\hat{\theta}) \approx 30,000$.
- Reject the model!
- Possible failures:

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y; \hat{\theta}) \approx 30,000$.
- Reject the model!
- Possible failures:
  - Error in the specification of A, r

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y; \hat{\theta}) \approx 30,000$.
- Reject the model!
- Possible failures:
  - Error in the specification of A, r
  - Lack of independence

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y; \hat{\theta}) \approx 30,000$.
- Reject the model!
- Possible failures:
  - Error in the specification of A, r
  - Lack of independence
  - super-Poisson variance in the counts

Posterior predictive checking

- Poisson noise + model problems makes exact non-negative solutions impossible.
- Use an estimate $\hat{g}$ and capture the discrepancy between y and $\hat{\theta} = A\hat{g} + r$.
- Consider χ^2 discrepancy.

$$X^2(y; \hat{\theta}) = \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}$$

- Fitting to real data: y with $n = 22464$.
- The best-fit non-negative image $\hat{g}$ was not an exact fit leading to the discrepancy between y and $\hat{\theta}$ to be $X^2(y; \hat{\theta}) \approx 30,000$.
- Reject the model!
- Possible failures:
  - Error in the specification of A, r
  - Lack of independence
  - super-Poisson variance in the counts
  - Error from discretizing the continuous g.

Posterior predictive checking

- The "critically bad" level of $X^2(y;\hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.

Posterior predictive checking

- The "critically bad" level of $X^2(y;\hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.
- We reassess and change the model.

Posterior predictive checking

- The "critically bad" level of $X^2(y; \hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.
- We reassess and change the model.
- We get a χ^2 discrepancy that is $22,000$ or even $20,000$.

Posterior predictive checking

- The "critically bad" level of $X^2(y;\hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.
- We reassess and change the model.
- We get a χ^2 discrepancy that is $22,000$ or even $20,000$.
- Should we just accept this new model?

Posterior predictive checking

- The "critically bad" level of $X^2(y;\hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.
- We reassess and change the model.
- We get a χ^2 discrepancy that is $22,000$ or even $20,000$.
- Should we just accept this new model?
- No.

Posterior predictive checking

- The "critically bad" level of $X^2(y;\hat{\theta})$ is $n + 2\sqrt{2n} \approx 23,000$.
- We reassess and change the model.
- We get a χ^2 discrepancy that is $22,000$ or even $20,000$.
- Should we just accept this new model?
- No.
- Be skeptical!

Another graph example

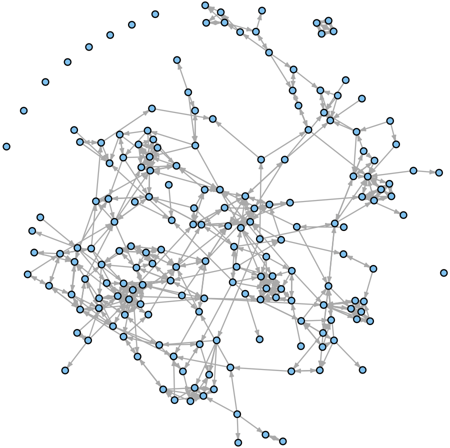
Relational Data

$$Y = \begin{pmatrix} - & Y_{12} & Y_{13} & \cdots & Y_{1m} \\ Y_{21} & - & & & \\ Y_{31} & & - & & \\ \vdots & & & - & \\ Y_{m1} & & & & - \end{pmatrix}$$

Y_{ij} is the relationship between nodes i and j.
When $Y_{ij} \in \{0, 1\}$ this is a sociomatrix.

High school in Adolescent Health data set

- 181 male respondents
- Each one nominated at most 5 friends
- There are people who nominated no one and were nominated by no one



Model - standard approach (SRM)

- The social relations model (SRM) was introduced by Warner, Kenny and Stoto (1979).
- Probit model with row and column effects:

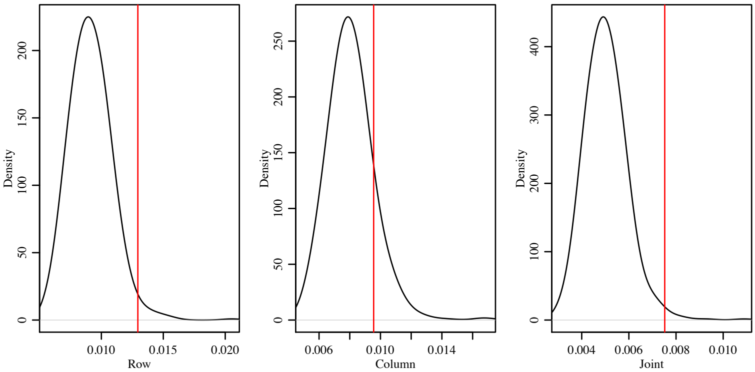
$$
\begin{aligned}
Y_{ij} &= \mathbf{1}_{Z_{ij}>0} \\
Z_{ij} &= \beta^t X_{ij} + a_i + b_j + \varepsilon_{ij} \\
(a_i\, b_i) &\stackrel{\text{iid}}{\sim} \text{normal}(0, \Sigma_{ab}) \\
\text{cor}(\varepsilon_{ij}, \varepsilon_{ji}) &= \rho
\end{aligned}
$$

- Note that $\text{cov}(Z_{ij}, Z_{kl}) = 0$ unless Z_{ij} and Z_{kl} are in the same column, same row, or are reciprocal.

Summary statistics

- Posterior predictive checks (PPC): at each iteration of the MCMC procedure we
  1. Sample from the full conditionals of $Z^{(s)}$.
  2. Simulate new data, $Y^{(s)}$.
  3. Calculate test statistics.
- Statistics to consider:
  - Binary row correlation: $t_{\text{row}}\left(Y^{(s)}\right) =$ average of $\text{cor}\left(Y^{(s)}_{i,-(i,j)}, Y^{(s)}_{j,-(i,j)}\right)^2$.
  - Binary column correlation: $t_{\text{col}}\left(Y^{(s)}\right) =$ average of $\text{cor}\left(Y^{(s)}_{-(i,j),i}, Y^{(s)}_{-(i,j),j}\right)^2$
  - Binary joint correlation: $t_{\text{joint}}\left(Y^{(s)}\right) =$ average of $\text{cor}\left(\left(Y^{(s)}_{i,-(i,j)}, Y^{(s)}_{-(i,j),i}\right), \left(Y^{(s)}_{j,-(i,j)}, Y^{(s)}_{-(i,j),j}\right)\right)^2$

Summary statistics



Observed statistic. Excess correlation is present.

A particular model for row and column covariance

- SRM is able to capture covariance within a row or within a column:

$$\mathrm{cov}(Z_{ij}, Z_{kl}) = \begin{cases} \sigma_a^2 & i = k, j \neq l \\ \sigma_b^2 & i \neq k, j = l \\ 0 & (i \neq k, j \neq l) \end{cases}$$

- We propose a matrix normal model that can capture correlation between Z_{ij} and Z_{kl} which are in different rows and columns.