Single parameter models

Dhafer Malouche

Outline

Probability distributions

Bernoulli Distribution

Bernoulli Distribution: Definition

- Discrete probability distribution for **binary outcomes**
- Random variable X takes values:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

- Probability mass function (PMF):

$$P(X = k) = p^k (1 - p)^{1-k} \quad \text{for } k \in \{0, 1\}$$

- Parameters: $p \in [0, 1]$ (probability of success)

Real-World Example

Login Attempt Success

Consider a website login system where:

- Each login attempt succeeds with probability $p = 0.3$
- Fails with probability 0.7

Let:

$$X = \begin{cases} 1 & \text{Successful login} \\ 0 & \text{Failed login} \end{cases}$$

This is a Bernoulli(p $=$ 0.3) random variable.

Simulation and Comparison in R


```r
# Parameters
n <- 1000
p <- 0.3

# Generate random samples
set.seed(123)
samples <- rbinom(n, 1, p)

# Calculate empirical probabilities
emp_probs <- table(samples)/n
```

Theoretical results:

- P(X=0) = 0.7
- P(X=1) = 0.3

Empirical results:

- P(X=0) = emp_probs[1]
- P(X=1) = emp_probs[2]

Visual Comparison

PMF Comparison:

```
# Theoretical PMF
theoretical <- c(0.7, 0.3)

# Plot comparison
barplot(rbind(theoretical, emp_probs),
        beside = TRUE,
        col = c("red", "blue"),
        main = "PMF Comparison")
```

Visual Comparison

CDF Comparison:

```
# Plot theoretical CDF
plot(stepfun(c(0,1), c(0,0.7,1)),
     main = "CDF Comparison",
     xlim = c(-0.5,1.5))

# Add empirical CDF
lines(ecdf(samples),
      col = 'blue', lty = 2)
```

Activity: Simulating Disease Transmission

Scenario: COVID-19 exposure at a school (N = 500 students)

Simulation Steps:

```r
# Set parameters:
set.seed(456)
# Transmission probability
p_infect <- 0.15
n_students <- 500
# Simulate infections:
infected <- rbinom(n_students, 1, p_infect)
# Calculate outcomes:
n_infected <- sum(infected)
attack_rate <- mean(infected)
# Visualize
barplot(c(Theoretical = p_infect,
          Observed = attack_rate),
        ylab = "Probability",
        col = c("red3", "dodgerblue"),
        main = "Infection Risk Comparison")
```

Immediate Tasks:

- Report number infected
- Compare red vs. blue bars
- Compute: $\frac{\text{Infected}}{\text{Total}}$
- Expected vs. observed cases

Activity: Analysis & Public Health Insights (Part 1)

Follow-up from Simulation Results

```
# Confidence interval for attack rate
prop.test(n_infected, n_students)

# Simulate new variant impact
p_new_variant <- 0.30  # 30% infection probability
infected_new <- rbinom(n_students, 1, p_new_variant)
```

Initial Questions:

- How would you interpret the 95% confidence interval?
- What does changing p_new_variant from 0.15 to 0.30 imply about:
  - Hospitalization rates?
  - Healthcare system capacity?

Key concepts introduced: Confidence intervals, parameter sensitivity

Activity: Analysis & Public Health Insights (Part 2)

Vaccine Efficacy (VE)

- Definition: Relative reduction in infection risk for vaccinated vs unvaccinated
- Formula:

$$VE = 1 - \frac{\text{Risk}_{\text{vaccinated}}}{\text{Risk}_{\text{unvaccinated}}}$$

- R implementation:

```
p_control <- 0.25   # Infection rate in control group
p_vax <- 0.05       # Infection rate in vaccinated group
ve <- 1 - (p_vax/p_control)  # Returns 0.8 (80% efficacy)
```

Sample Size Calculation

- Theory: For 95% CI with margin of error e: We would like to know the needed sample if we would a 95% CI mean $\pm$ 2%

$$n = \frac{Z_{\alpha/2}^2 \cdot p(1-p)}{e^2}$$

- R implementation:

```
p <- 0.5    # Conservative estimate
e <- 0.02   # Desired precision
n <- (qnorm(0.975)^2 * p*(1-p)) / e^2
```

Binomial Distribution

Binomial Distribution: Definition

- Discrete distribution for **success counts in n trials**
- $X \sim \text{Bin}(n, p)$ where:
  - n: Number of independent trials
  - p: Success probability per trial
- PMF:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \ldots, n$$

- Expectation: $\mathbb{E}[X] = np$
  Variance: $\text{Var}(X) = np(1 - p)$

Epidemiological Example

Vaccine Efficacy Trial

In a COVID-19 vaccine trial with 100 participants:

- Each participant has 5% infection risk (placebo group)
- Let X = number of infections in the group
- $X \sim \text{Bin}(n = 100, p = 0.05)$

Probability questions:

- $P(X \geq 10)$ - Extreme outbreak risk
- $P(5 \leq X \leq 15)$ - Expected range

Simulation and Comparison in R

```
# Parameters
n_trials <- 1000
n <- 100    # Participants
p <- 0.05   # Infection risk

# Simulate outbreaks
set.seed(123)
infections <- rbinom(n_trials, n, p)

# Theoretical PMF
k <- 0:n
pmf <- dbinom(k, n, p)
```

Comparison checks:

- Empirical mean vs np
- Sample variance vs $np(1-p)$
- Histogram shape vs PMF

```
mean(infections)
# Should be ~5
var(infections)
# Should be ~4.75
```

Visual Comparison

- PMF Comparison:

```r
hist(infections, freq=FALSE,
     main = "Infection Distribution",
     xlab = "Number of Cases")
lines(k, pmf, col="red", lwd=2)
```

- CDF Comparison:

```r
plot(ecdf(infections),
     main = "CDF Comparison")
lines(k, pbinom(k, n, p),
      col="red", type="s")
```

Poisson Distribution

Poisson Distribution: Definition

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

- Discrete distribution for **event counts in fixed interval/area**
- $X \sim \text{Pois}(\lambda)$ where:
  - λ: Average rate (events per unit)
  - Constant event risk, independent occurrences
- PMF:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad k = 0, 1, 2, \ldots$$

- Expectation/Variance: $\mathbb{E}[X] = \text{Var}(X) = \lambda$

Environmental Science Example

Microplastic Pollution Monitoring

Water quality researchers study a river:

- Average of 4 microplastic particles per liter ($\lambda = 4$)
- $X =$ number of particles in 1L sample
- $X \sim \text{Pois}(4)$

Key questions:

- $P(X \geq 8)$ - Extreme pollution probability
- $P(X \leq 2)$ - Compliance with safety standards
- 95% prediction interval for particle counts

Simulation and Comparison in R


```r
# Parameters
lambda <- 4
n_samples <- 1000

# Simulate water samples
set.seed(123)
particles <- rpois(n_samples, lambda)

# Theoretical PMF
k <- 0:15
pmf <- dpois(k, lambda)
```

Validation checks:

- Empirical mean: `mean(particles)`
- Sample variance: `var(particles)`
- Zero-inflation test: `mean(particles==0)`

```r
# Expected
# vs observed zeros
exp(-lambda)
# Theoretical P(X=0)
mean(particles == 0)
```

Visual Comparison

- PMF Comparison:

```
hist(particles, freq=FALSE, breaks=0:15,
     main = "Microplastic Particles/Liter",
     xlab = "Count")
points(k, pmf, col="red", pch=19)
```

- Time Series:

```
plot(particles[1:50], type="b",
     main = "Particle Count Time Series",
     xlab = "Sample ID", ylab = "Count",
     col = "darkgreen")
abline(h = lambda, col="red", lty=2)
```

Activity: Pollution Hotspot Analysis

College of Arts and Sciences
QATAR UNIVERSITY

Scenario: Comparing two river sites

Simulation Task:

```r
# Downstream (polluted)
lambda_ds <- 8
# Upstream (reference)
lambda_us <- 3

samples <- 500
set.seed(456)
ds_counts <- rpois(samples, lambda_ds)
us_counts <- rpois(samples, lambda_us)
```

Analysis:

- Test $\lambda_{ds} > \lambda_{us}$
- Compute $P(\text{ds} \geq 10)$
- Visualize PMF comparisons
- Estimate pollution ratio
- Check dispersion (var/mean ratio)

Negative Binomial Distribution

Negative Binomial Distribution: Definition (Traditional)

- Discrete distribution for the number of trials needed to achieve r successes.
- Assumes independent Bernoulli trials with success probability p.
- PMF:

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \ldots$$

- Moments:
  - Mean: $\mathbb{E}[X] = \frac{r}{p}$
  - Variance: $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$

Customer Acquisition Costs

A SaaS company models trial-to-paid conversions:

- Each website visit is a trial with conversion probability $p = \frac{1}{3}$.
- A conversion requires one successful visit ($r = 1$).
- X = number of visits until conversion.
- $X \sim \text{NB}(r = 1, p = \frac{1}{3})$ (geometric distribution).

Key business questions:

- What is the probability of conversion in ≤ 2 visits?
- What is the 95th percentile of required visits?
- How does this inform customer acquisition cost modeling?

Simulation and Comparison in R

```
# Parameters
r <- 1          # Conversion requires 1 success
p <- 1/3        # Conversion probability per visit
n <- 1000

# Simulate number of visits until conversion
set.seed(2023)
failures <- rnbinom(n, size = r, prob = p)
visits <- failures + r  # Total visits (failures + 1 success)

# Theoretical PMF for number of visits
k <- 1:15
pmf <- p * (1 - p)^(k - 1)

# Empirical dispersion check
dispersion <- var(visits) / mean(visits)
dispersion  # Should be > 1 indicating overdispersion
```

Visual Comparison

- PMF Comparison:

```
hist(visits, freq=FALSE, breaks=0:20,
     main = "Visits Until Conversion",
     xlab = "Website Visits")
lines(k, pmf, col="red", type="h")
```

- QQ-Plot:

```
# Compare to Poisson
qqplot(rpois(1000, mu), visits,
       main = "NB vs Poisson QQ-Plot")
abline(0,1, col="red")
```

Normal Distribution

Normal Distribution: Definition

- Continuous distribution for **symmetric, bell-shaped data**
- $X \sim N(\mu, \sigma^2)$ where:
  - μ: Mean (location parameter)
  - σ: Standard deviation (scale parameter)
- Probability density function (PDF):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Properties:
  - Symmetric about μ
  - 68-95-99.7 rule for standard normal

Psychology Example

IQ Score Distribution

Standardized intelligence testing:

- IQ scores follow $N(100, 15^2)$
- Population parameters:
  - Mean $\mu = 100$
  - SD $\sigma = 15$
- Diagnostic thresholds:
  - Gifted: > 130 (top 2.3%)
  - Intellectual disability: < 70 (bottom 2.3%)

Simulation and Comparison in R


```r
# Parameters
mu <- 100
sigma <- 15
n <- 1000

# Generate scores
set.seed(123)
iq_scores <- rnorm(n, mu, sigma)

# Theoretical PDF
x <- seq(55, 145, length=100)
pdf <- dnorm(x, mu, sigma)
```

Validation checks:

- Empirical mean/SD vs parameters
- Shapiro-Wilk normality test
- Check 68-95-99.7 rule

```r
mean(iq_scores)
# Should be ~100
sd(iq_scores)
# Should be ~15
shapiro.test(iq_scores)
```

- Density Comparison:

```
hist(iq_scores, freq=FALSE, breaks=20,
     main = "IQ Score Distribution",
     xlab = "IQ")
lines(x, pdf, col="red", lwd=2)
```

- QQ-Plot:

```
qqnorm(iq_scores, main = "Normal Q-Q Plot")
qqline(iq_scores, col="red")
```

Exponential Distribution

Exponential Distribution: Definition

- Continuous distribution for **time between events in Poisson process**
- $X \sim \text{Exp}(\lambda)$ where:
  - λ: Rate parameter (events per unit time)
  - Mean time between events: $\mu = 1/\lambda$
- PDF:

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

- Key property: Memoryless

$$P(X > s + t | X > s) = P(X > t)$$

Cybersecurity Example

Network Intrusion Detection

A corporate network experiences:

- Average of 1 intrusion every 10 days ($\mu = 10$)
- $\lambda = 0.1$ intrusions/day
- $X =$ days between intrusions
- $X \sim \text{Exp}(0.1)$

Security questions:

- $P(\text{Next intrusion} < 24\text{hrs}) = ?$
- 90th percentile of safe period
- Probability of > 30 days without intrusion

Simulation and Comparison in R


```r
# Parameters
lambda <- 0.1  # Rate
mu <- 1/lambda # Mean = 10
n <- 1000

# Simulate inter-arrival times
set.seed(123)
durations <- rexp(n, lambda)

# Theoretical PDF
x <- seq(0, 30, length=100)
pdf <- dexp(x, lambda)
```

Validation checks:

- Empirical mean vs $1/\lambda$
- Variance vs $1/\lambda^2$
- Memoryless property test

```r
mean(durations)
# Should be ~10
var(durations)
# Should be ~100
```

Visual Comparison

- Density Comparison:

```
hist(durations, freq=FALSE, breaks=30,
     main = "Time Between Intrusions",
     xlab = "Days")
lines(x, pdf, col="red", lwd=2)
```

- Survival Function:

```
plot(ecdf(durations),
     main = "CDF Comparison")
lines(x, pexp(x, lambda),
      col="red", lwd=2)
```

Gamma Distribution

Gamma Distribution: Definition

- Continuous distribution for **positive-valued, skewed data**
- $X \sim \Gamma(k, \theta)$ where:
  - k: Shape parameter (controls skewness)
  - θ: Scale parameter
- PDF:

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \quad x > 0$$

- Properties:
  - $\mathbb{E}[X] = k\theta$, $\mathrm{Var}(X) = k\theta^2$
  - Generalizes Exponential ($k = 1$) and χ^2 distributions

Healthcare Example

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

Hospital Length of Stay

Modeling patient stays in a surgical ward:

- Average stay duration: 5 days ($k = 2, \theta = 2.5$)
- X = number of days hospitalized
- $X \sim \Gamma(2, 2.5)$

Clinical applications:

- $P(\text{Stay} > 10 \text{ days})$ - Identify long-stay patients
- 75th percentile for resource planning
- Compare recovery times between procedures
- Model healthcare costs associated with stays

Why Gamma?

Simulation and Comparison in R

```r
# Parameters
shape <- 2
scale <- 2.5
n <- 1000

# Simulate hospital stays
set.seed(123)
stays <- rgamma(n, shape=shape, scale=scale)

# Theoretical PDF
x <- seq(0, 20, length=100)
pdf <- dgamma(x, shape=shape, scale=scale)
```

Validation checks:

- Empirical mean vs $k\theta$
- Variance vs $k\theta^2$
- Skewness assessment

```r
mean(stays)
# Should be ~5
var(stays)
# Should be ~12.5
moments::skewness(stays)
# ~1.41
```

Visual Comparison

- Density Comparison:

```
hist(stays, freq=FALSE, breaks=30,
     main = "Hospital Stay Duration",
     xlab = "Days", col="lightblue")
lines(x, pdf, col="maroon", lwd=2)
legend("topright",
        legend=c("Theory","Empirical"),
       col=c("maroon","lightblue"), lwd=2)
```

- CDF Comparison:

```
plot(ecdf(stays),
    main="Empirical vs Theoretical CDF")
lines(x, pgamma(x, shape=shape, scale=scale),
      col="darkgreen", lwd=2)
```

Beta Distribution

Beta Distribution: Definition

- Continuous distribution for **probabilities/proportions** (0-1)
- $X \sim \text{Beta}(\alpha, \beta)$ where:
  - α: Shape 1 (successes + 1)
  - β: Shape 2 (failures + 1)
- PDF:

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad 0 \leq x \leq 1$$

- Properties:
  - $\mathbb{E}[X] = \dfrac{\alpha}{\alpha+\beta}$
  - $\text{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
  - Generalizes Uniform ($\alpha = \beta = 1$)

Marketing Example

A/B Test Conversion Rates

Comparing website versions:

- Version A: 45 conversions / 1000 visits
- Version B: 65 conversions / 1000 visits
- Model conversion rates as:
  - $r_A \sim \text{Beta}(46, 956)$
  - $r_B \sim \text{Beta}(66, 936)$

Business questions:

- $P(r_B > r_A)$ - Version superiority probability
- 95% credible intervals for rates
- Minimum detectable effect size

Why Beta?

- Natural for bounded probabilities
- Conjugate prior for Binomial
- Flexible uncertainty representation

Simulation and Comparison in R

```r
# Parameters
alpha <- 46
beta <- 956
n <- 10000

# Simulate conversion rates
set.seed(123)
rates <- rbeta(n, alpha, beta)

# Theoretical PDF
x <- seq(0, 0.1, length=100)
pdf <- dbeta(x, alpha, beta)
```

Validation checks:

- Empirical mean vs $\alpha/(\alpha+\beta)$
- Variance comparison
- Uniform check when $\alpha = \beta = 1$

```r
mean(rates)
# Should be ~0.046
quantile(rates, c(0.025,0.975))
```

Visual Comparison

College of Arts and Sciences
QATAR UNIVERSITY

- Density Comparison:

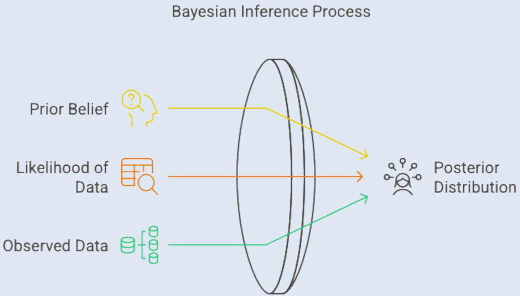
```
hist(rates, freq=FALSE, breaks=50,
     main = "Conversion Rate Distribution",
     xlab = "Conversion Rate", col="skyblue")
lines(x, pdf, col="darkred", lwd=2)
```

- CDF Comparison:

```
plot(ecdf(rates), main="Beta CDF Comparison")
lines(x, pbeta(x, alpha, beta),
      col="darkgreen", lwd=2)
```

Bayesian single parameter Models

Single Parameter Mode



$$\underbrace{p(\theta \mid \text{Data})}_{\text{posterior}} \propto \underbrace{L(\text{Data} \mid \theta)}_{\text{likelihood}} \times \underbrace{p(\theta)}_{\text{prior}}$$

Bernoulli Model

Example: Influence of Prior Beliefs on Bayesian Updating

- Scenario: Estimating the probability of success for a new treatment based on observed trial results.
- Observations: Suppose you observe 10 trials of a new treatment, where 6 trials are successful.
- Prior Beliefs:
  - **Non-Informative Prior**: We start with no prior information, reflecting a neutral stance on the treatment's effectiveness. This prior does not influence the likelihood gained from the observed data.
  - **Weakly Informative Prior**: We have some previous experience or expert opinion suggesting the treatment's success rate is around 50

Choosing the Prior Distribution

- **Prior**: Represents initial beliefs about p (probability of success).
- Common choices:
  - **Conjugate Prior (Beta)**:

$$p \sim \text{Beta}(\alpha, \beta)$$

    - α: Prior successes $+ 1$
    - β: Prior failures $+ 1$
  - **Non-Informative Prior**: $\text{Beta}(1, 1)$ (uniform over $[0, 1]$)
  - **Weakly Informative Prior**: $\text{Beta}(2, 2)$ (gentle nudge toward 0.5)

Posterior Distribution

Given observed data $y = (y_1, y_2, \ldots, y_n)$, the posterior is:

$$p \mid y \sim \text{Beta}(\alpha + \text{successes}, \beta + \text{failures})$$

- **Successes**: $\sum_{i=1}^{n} y_i$
- **Failures**: $n - \sum_{i=1}^{n} y_i$

R Example: Analytical Solution


```r
# Data
successes <- 7
failures <- 3

# Priors
prior_uniform <- c(1, 1)          # Non-informative
prior_weak <- c(2, 2)             # Weakly informative
prior_informative <- c(5, 5)      # Informative

# Posteriors
posterior_uniform <- c(prior_uniform[1] + successes,
                        prior_uniform[2] + failures)
posterior_weak <- c(prior_weak[1] + successes,
                    prior_weak[2] + failures)
posterior_informative <- c(prior_informative[1] + successes,
                    prior_informative[2] + failures)
```

R Example: Analytical Solution

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

```
# Plot
curve(dbeta(x, posterior_uniform[1], posterior_uniform[2]),
      xlab="p", ylab="Density", col="blue",
      lwd=2, ylim=c(0, 4))
curve(dbeta(x, posterior_weak[1],
     posterior_weak[2]),
      col="red", lwd=2, add=TRUE)
curve(dbeta(x, posterior_informative[1],
      posterior_informative[2]),
      col="green", lwd=2, add=TRUE)
legend("topright",
       legend=c("Uniform", "Weak", "Informative"),
       col=c("blue", "red", "green"), lwd=2)
```

R Example: Using rstan

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

Stan Code (bernoulli_model.stan):

```
data {
  int<lower=0> N;          // Number of trials
  int<lower=0, upper=1> y[N];  // Binary outcomes
}
parameters {
  real<lower=0, upper=1> p;  // Probability of success
}
model {
  p ~ beta(1, 1);          // Uniform prior
  y ~ bernoulli(p);        // Likelihood
}
```

R Example: Using rstan

R Code to Run Stan:

```
library(rstan)
y <- c(1, 0, 1, 1, 0, 1, 1, 1, 0, 1)
stan_data <- list(N = length(y), y = y)
fit <- stan(file = "bernoulli_model.stan",
       data = stan_data,
      iter = 2000, chains = 4)
print(fit)  # Summary of posterior
plot(fit)   # Visualize posterior distribution
```

Activity: Vaccine Efficacy

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

Scenario: A new vaccine is tested on 100 individuals. 15 develop side effects.

Task:

- Define a Beta prior for side effect probability p.
- Compute the posterior distribution.
- Compare results for:
  - Non-informative prior ($\text{Beta}(1, 1)$)
  - Weakly informative prior ($\text{Beta}(2, 2)$)
  - Informative prior ($\text{Beta}(5, 95)$)
- Use `rstan` to fit the model and compare results.

Questions:

- What is $P(p > 0.2)$ under each prior?
- How does the choice of prior affect the posterior?
- What sample size would reduce prior influence?

Normal Model

Case 1: Unknown Mean, Known Variance

كلية الآداب والعلوم
College of Arts and Sciences
QATAR UNIVERSITY جامعة قطر

- **Model**: $y \sim N(\mu, \sigma^2)$, where σ^2 is known.
- **Conjugate Prior**: Normal distribution for μ:

$$\mu \sim N(\mu_0, \tau_0^2)$$

  - μ_0: Prior mean
  - τ_0^2: Prior variance

- **Posterior**:

$$\mu \mid y \sim N\left(\mu_n, \tau_n^2\right)$$

  where:

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

R Example: Unknown Mean, Known Variance


```r
# Data
y <- c(5.1, 5.5, 4.9, 5.3, 5.7)  # Sample data
n <- length(y)
sigma2 <- 0.5  # Known variance

# Prior
mu0 <- 5.0  # Prior mean
tau02 <- 1.0  # Prior variance

# Posterior
mu_n <- (mu0/tau02 + n*mean(y)/sigma2) / (1/tau02 + n/sigma2)
tau_n2 <- 1 / (1/tau02 + n/sigma2)

# Plot
curve(dnorm(x, mu0, sqrt(tau02)), xlim=c(4, 6), ylab="Density",
      col="blue", lwd=2, main="Prior vs Posterior")
curve(dnorm(x, mu_n, sqrt(tau_n2)), col="red", lwd=2, add=TRUE)
legend("topright", legend=c("Prior", "Posterior"),
       col=c("blue", "red"), lwd=2)
```

Case 2: Unknown Mean and Variance

- **Model**: $y \sim N(\mu, \sigma^2)$, both μ and σ^2 unknown.
- **Conjugate Prior**: Normal-Inverse-Gamma (NIG):

$$\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0), \quad \sigma^2 \sim \text{Inv-Gamma}(\alpha_0, \beta_0)$$

- μ_0: Prior mean
- κ_0: Prior precision scaling
- α_0, β_0: Shape and scale for σ^2

- **Posterior**:

$$\mu \mid \sigma^2, y \sim N(\mu_n, \sigma^2/\kappa_n), \quad \sigma^2 \mid y \sim \text{Inv-Gamma}(\alpha_n, \beta_n)$$

where:

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \quad \kappa_n = \kappa_0 + n, \quad \alpha_n = \alpha_0 + \frac{n}{2},$$

$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{\kappa_0 n(\bar{y} - \mu_0)^2}{2(\kappa_0 + n)}$$

R Example: Unknown Mean and Variance


```r
# Data
y <- c(5.1, 5.5, 4.9, 5.3, 5.7)  # Sample data
n <- length(y)

# Prior
mu0 <- 5.0  # Prior mean
kappa0 <- 1  # Prior precision scaling
alpha0 <- 2  # Prior shape for sigma^2
beta0 <- 1   # Prior scale for sigma^2

# Posterior
mu_n <- (kappa0 * mu0 + n * mean(y)) / (kappa0 + n)
kappa_n <- kappa0 + n
alpha_n <- alpha0 + n/2
beta_n <- beta0 + 0.5*sum((y - mean(y))^2) +
          (kappa0 *n*(mean(y) - mu0)^2)/(2*(kappa0 + n))
```

R Example: Unknown Mean and Variance

```
# Plot posterior for mu
curve(dnorm(x, mu_n, sqrt(beta_n / (alpha_n * kappa_n))),
      xlim=c(4, 6), ylab="Density", col="red", lwd=2,
      main="Posterior for Mean")
```

Using `rstan` for Unknown Mean and Variance

Stan Code (`normal_model.stan`):

```
data {
  int<lower=0> N;          // Number of observations
  real y[N];               // Data
}
parameters {
  real mu;                 // Mean
  real<lower=0> sigma2; // Variance
}
model {
  mu ~ normal(5.0, sqrt(sigma2));  // Prior for mu
  sigma2 ~ inv_gamma(2, 1);        // Prior for sigma2
  y ~ normal(mu, sqrt(sigma2));    // Likelihood
}
```

Using rstan for Unknown Mean and Variance

R Code to Run Stan:

```
library(rstan)
y <- c(5.1, 5.5, 4.9, 5.3, 5.7)  # Sample data
stan_data <- list(N = length(y), y = y)
fit <- stan(file = "normal_model.stan", data = stan_data, ite
print(fit)  # Summary of posterior
plot(fit)   # Visualize posterior distributions
```

Activity: Climate Change Analysis

Scenario: Analyze annual temperature anomalies (in °C) for a city:

$$y = \{0.9, 1.1, 1.3, 1.0, 1.2, 1.4, 1.1, 1.3, 1.5, 1.2\}$$

Task:

- Assume $\sigma^2 = 0.1$ is known. Use a Normal$(1.0, 0.5)$ prior for μ. Compute the posterior.
- Assume both μ and σ^2 are unknown. Use a NIG$(1.0, 1, 2, 1)$ prior. Compute the posterior.
- Compare results using `rstan`.

Questions:

- What is $P(\mu > 1.2)$ under each model?
- How does prior choice affect results?
- What sample size reduces prior influence?

Poisson Model

Introduction to Poisson Model

- The Poisson distribution is often used to model the number of events occurring within a fixed period of time.
- **Model Assumption**: Events occur independently at a constant rate.
- Common applications: Counting the number of occurrences of events (e.g., arrival of customers, mutation occurrences in a DNA sequence).

Bayesian Poisson Model

- In Bayesian inference, we combine prior beliefs about a parameter with the likelihood of observed data.
- **Likelihood**: Assuming y events observed follows $y \sim \text{Poisson}(\lambda)$.
- **Prior for λ**: Typically a Gamma distribution due to its conjugacy, $\lambda \sim \text{Gamma}(a, b)$.
  - A non-informative prior can be implemented by setting a and b very close to zero (e.g., 0.001). This results in a very flat distribution, indicating high uncertainty and allowing the data to have a stronger influence on the posterior.
  - The flatness ensures that the prior does not overly constrain the likelihood, providing minimal initial information about the parameter.

Poisson Model in RStan

College of Arts and Sciences
QATAR UNIVERSITY

Stan Code (poisson_model.stan):

```
data {
  int<lower=0> N;         // Number of events observed
  int y[N];               // Observed counts
}
parameters {
  real<lower=0> lambda; // Rate parameter of Poisson
}
model {
  // Non-informative prior: lambda ~ gamma(0.001, 0.001)
  // Informative prior: lambda ~ gamma(9, 0.5)
  lambda ~ gamma(9, 0.5);      // Change as needed
  y ~ poisson(lambda);
}
```

Poisson Model in RStan

R Code to Run Stan:

```
library(rstan)
data_obs <- c(3, 2, 1, 5, 4)  // Example data
stan_data <- list(N = length(data_obs), y = data_obs)
fit <- stan(file = "poisson_model.stan", data = stan_data,
            iter = 2000, chains = 4)
print(fit)  // Summary of posterior
plot(fit)   // Visualize posterior distribution
```