TDAB01 Sannolikhetslära och Statistik

Jose M. Peña
IDA, Linköpings Universitet

Föreläsning 10

Översikt

- **Bayesiansk inferens**
- **Bernoullimodell med beta prior**
- **Normalmodell med normal prior**
- **Multinomialmodell med Dirichlet prior**
- **Bayesiansk estimator, konfidensintervall och hypotestest**

Frekventistisk inferens

- Hittills på kursen: **Frekventistisk inferens**.
  - **Parametrar θ är fixa** (icke slumpmässiga) storheter.
  - **Data är slumpvariabler**: $f(x_1, \ldots, x_n|\theta)$.
- Frekventistisk inferens: Hur en **metod** beter sig över **upprepade stickprov** från populationen.
- **Samplingfördelningar** är i fokus, dvs vilka värden kan min estimator förväntas anta för olika stickprov ?
- **Väntevärderiktighet**: Min skattningsmetod kommer att vara korrekt i genomsnitt, dvs sett över alla möjliga stickprov.
- **Konfidensintervall**: Min intervallskattningsmetod kommer att täcka det sanna parametervärdet θ i 95% av alla möjliga stickprov från populationen.
- **Hypotestest**: Min testmetod kommer bara att dra fel slutsats i 5% av alla stickprov om nollhypotesen är sann.

Subjektiva sannolikheter

- Du **vet inte** värdet på en populationsparameter θ. Du är **osäker** om θ. Påståendet $\boldsymbol{P}(\theta \leq 2)$ är meningsfullt.
- **Det är osäkerheten som är relevant**. Om θ är en fix, konstant, storhet eller ej spelar ingen roll.
- Jag vet inte 10:e decimalen av π. Då kan jag säga

$$P(10:e \text{ decimal av } \pi = 9) = 1/10$$

- Det är **min** osäkerhet som spelar roll. Du kanske vet 10:e decimalen av π. För mig är π osäker och jag kan prata om sannolikhetsfördelningen för 10:e decimalen av π.
- Sannolikheter är ett **subjektivt** mått på personlig **grad av tilltro**.
- **Bayesiansk statistik** bygger på ett subjektivt sannolikhetsbegrepp.

Thomas Bayes 1701-1761

Bayesiansk inferens

- Bernoullimodellen: $X_1, \ldots, X_n | \theta \sim Bernoulli(\theta)$. T ex slantsingling.
- Sannolikheten θ för krona är okänd.
- Innan vi har börjat singla slant beskriver jag min osäkerhet om θ med min **apriorifördelning**: $\pi(\theta)$.
- **A priori** $=$ **före** (före jag har observerat data).
- Antag nu att vi har observerat ett antal slantsinglingar $X_1 = x_1, \ldots, X_n = x_n$. T ex $0, 0, 1, 1, 0$.
- Hur bör vi **uppdatera** vår apriorifördelning med denna datainformation ? Hur **lär** vi oss från data ?.
- **Aposteriorifördelning**: $\pi(\theta | x_1, ..., x_n)$.
- **A posterior** $=$ **efter** (efter jag har observerat data).
- Bayesiansk inferens **betingar på observerade data**. Då $P(\text{Okänt} \mid \text{Känt})$.

Bayes sats uppdaterar prior till posterior: Diskreta fallet

- Antag att θ bara kan anta värdena $0.1, 0.2, \ldots, 0.9$ (diskretisering).
- Kom ihåg **Bayes sats** för händelser A och B:

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

- Låt t ex $A = \{\theta = 0.1\}$ och $B = \{\mathrm{X} = \mathrm{x}\}$.
- Bayes sats ger **posteriorfördelningen**:

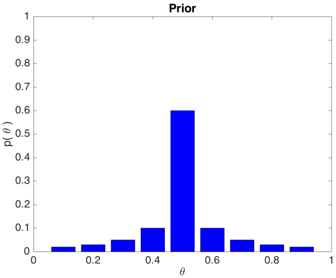
$$\boldsymbol{P}(\theta = 0.1|\mathrm{x}) = \frac{P(\mathrm{x}|\theta = 0.1)\boldsymbol{P}(\theta = 0.1)}{P(\mathrm{x})}$$

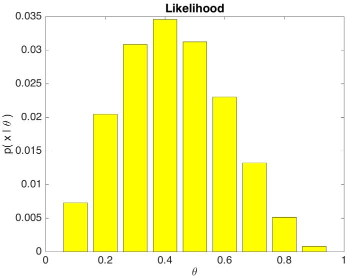
där satsen om total sannolikhet ger

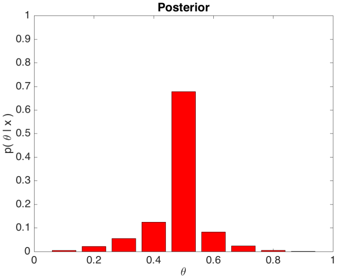
$$P(\mathrm{x}) = P(\mathrm{x}|\theta = 0.1)\boldsymbol{P}(\theta = 0.1) + \cdots + P(\mathrm{x}|\theta = 0.9)\boldsymbol{P}(\theta = 0.9)$$

- **Prior**: $\boldsymbol{P}(\theta = 0.1)$
- **Likelihood**: $P(\mathrm{x}|\theta = 0.1)$
- **Posterior**: $\boldsymbol{P}(\theta = 0.1|\mathrm{x})$

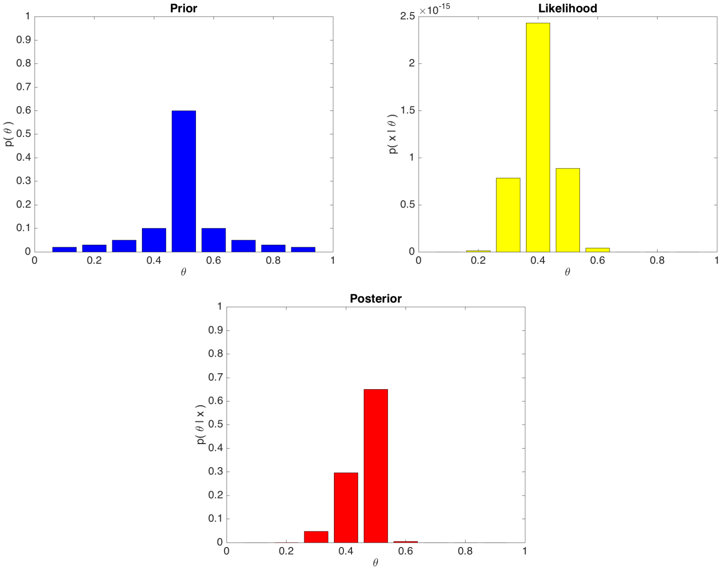
Bernoullimodell med s=2 och f=3



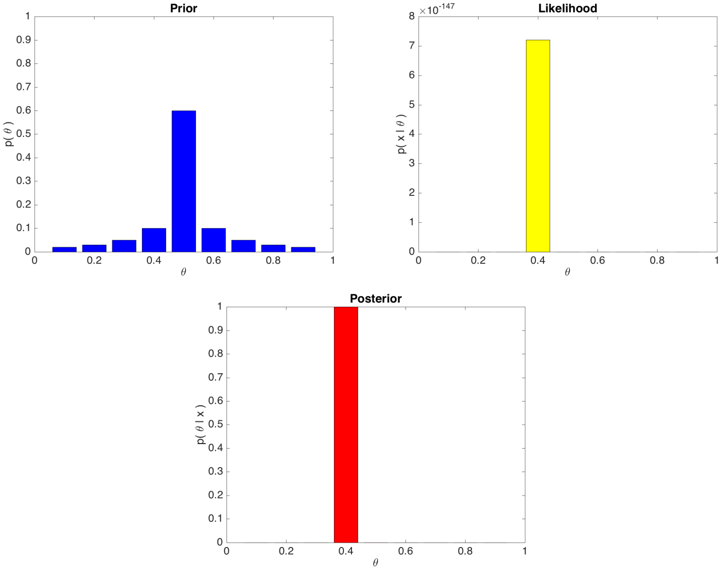




Bernoullimodell med s=20 och f=30



Bernoullimodell med s=200 och f=300



Bayes sats uppdaterar prior till posterior: Kontinuerliga fallet

- Diskretisering så att $\theta \in \{\theta_1, \theta_2, \ldots, \theta_K\}$

$$\boldsymbol{P}(\theta = \theta_i | \mathrm{x}) = \frac{P(\mathrm{x}|\theta = \theta_i)\boldsymbol{P}(\theta = \theta_i)}{\sum_{j=1}^{K} P(\mathrm{x}|\theta = \theta_j)\boldsymbol{P}(\theta = \theta_j)}$$

- Finare och finare grid $(\theta_{i+1} - \theta_i \to 0)$ ger

$$f(\theta|\mathrm{x}) = \frac{P(\mathrm{x}|\theta)f(\theta)}{\int P(\mathrm{x}|\theta)f(\theta)d\theta}$$

- **Bayes sats** för **kontinuerlig** parameter θ

$$\pi(\theta|\mathrm{x}) = \frac{f(\mathrm{x}|\theta)\pi(\theta)}{\int f(\mathrm{x}|\theta)\pi(\theta)d\theta}$$

  - **Prior**: $\pi(\theta)$
  - **Likelihood**: $f(\mathrm{x}|\theta)$
  - **Posterior**: $\pi(\theta|\mathrm{x})$

Subjektivitet och objektivitet

- $\pi(\theta)$ är en **subjektiv** fördelning som varierar från person till person baserat på erfarenhet, osv.
- **Hur vi lär oss från data**, dvs uppdaterar från prior till posterior, bestäms av Bayes sats.
- **Uppdateringmekanismen är objektiv** (matematik).
- Resultat: När $n \to \infty$ (**stora datamängder**) kommer alla personers posteriors att konvergera till samma fördelning. Objektivitet genom **subjektivt konsensus**.
- Vid rapportering av resultat kan man använda **icke-informativa** apriorifördelningar (dvs svag information) eller priorinformation som är lättförståelig.
- Maskininlärning: Mycket vanligt med aprioriinformation av typen "Jag tror att den okända funktionen är **mjuk**, men jag vet inte mycket mer om den exakta funktionsformen".

Bernoullimodell med beta prior

- Bernoullimodellen: $X_1, \ldots, X_n|\theta \sim Bernoulli(\theta)$. **Likelihood:** $\theta^s(1-\theta)^f$.
- $\theta \in [0,1]$. Lämplig **prior**: $\theta \sim Beta(\alpha, \beta)$, dvs

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- **Posterior:**

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} = \frac{\theta^s(1-\theta)^f\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int \theta^s(1-\theta)^f\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta}$$

$$= \frac{\theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}}{\int \theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}d\theta} = c \cdot \theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}$$

där $c = 1/\int \theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}d\theta$ är en konstant (beror inte på θ).

- En täthet på formen $c \cdot \theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}$ känns igen som en $Beta(\alpha+s, \beta+f)$:

$$\pi(\theta|x) = \frac{1}{B(\alpha+s,\beta+f)}\theta^{\alpha+s-1}(1-\theta)^{\beta+f-1}$$

Bayes sats på proportionell form

- Notera att vi aldrig behövde räkna ut nämnaren i Bayes sats, $\int f(x|\theta)\pi(\theta)d\theta$. Vi kände igen betafördelningen ändå.
- Tätheter måste integrera till ett. Proportionalitetskonstanter kan vi "strunta i".
- Enklare form av Bayes sats:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

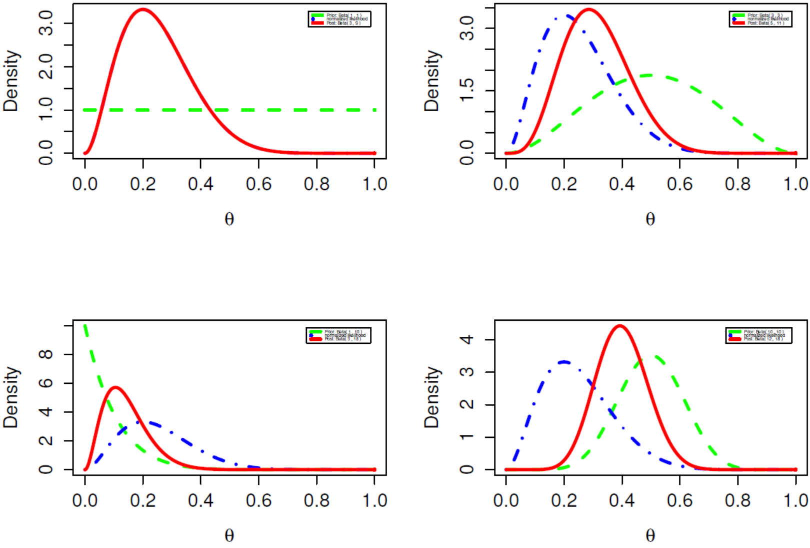
$$\text{Posterior} \ \propto \ \text{Likelihood} \ \times \ \text{Prior}$$

Exempel: Spam data

- George har gått igenom 4601 mejl, och 1813 av dessa var spam.
- **Modell**: Låt $x_i = 1$ om det i:te mejlet var spam. Antag $x_i|\theta \overset{iid}{\sim} Bernoulli(\theta)$.
- **Prior**: $\theta \sim Beta(\alpha, \beta)$.
- **Posterior**: $\theta|\mathrm{x} \sim Beta(\alpha + 1813, \beta + 2788)$.

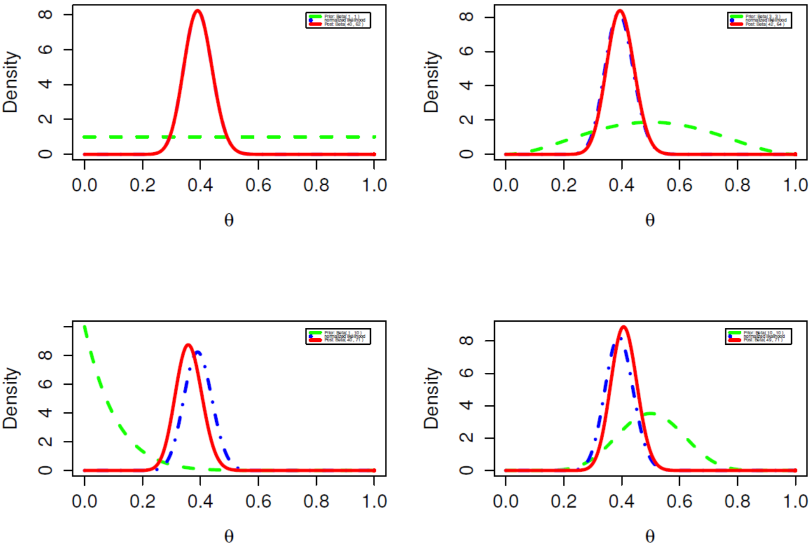
Exempel: Spam data (n=10) med fyra olika priors

- Grön: Prior. Blå: Normaliserad likelihood. Röd: Posterior.



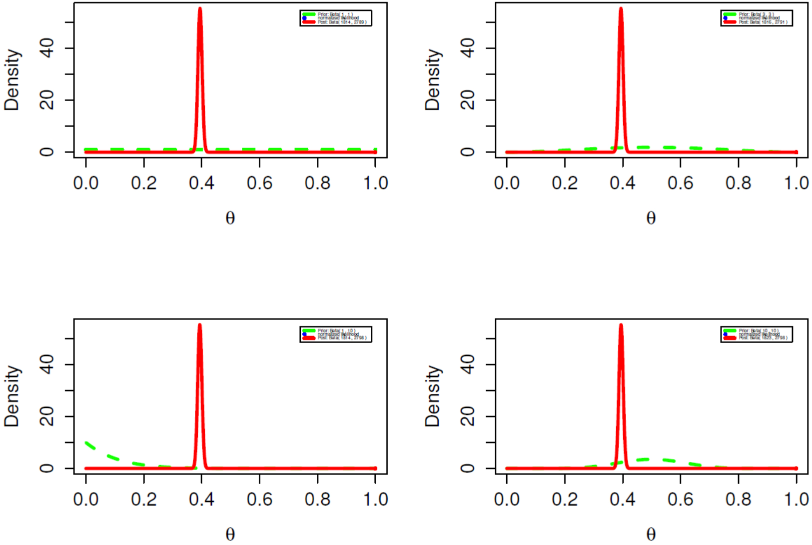
Exempel: Spam data (n=100) med fyra olika priors

- Grön: Prior. Blå: Normaliserad likelihood. Röd: Posterior.



Exempel: Spam data (n=4601) med fyra olika priors

- Grön: Prior. Blå: Normaliserad likelihood. Röd: Posterior.



Normalmodell med normal prior

- **Modell**: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ med σ^2 **känd**.
- **Prior**:

$$\theta \sim N(\mu, \tau^2)$$

- **Posterior**:

$$\begin{aligned} P(\theta|x_1, \ldots, x_n) &\propto P(x_1, \ldots, x_n|\theta, \sigma^2)p(\theta) \\ &\propto N(\theta|\mu_x, \tau_x^2) \end{aligned}$$

  där

$$\frac{1}{\tau_x^2} = \frac{n}{\sigma^2} + \frac{1}{\tau^2}$$

$$\mu_x = w\bar{x} + (1 - w)\mu$$

  och

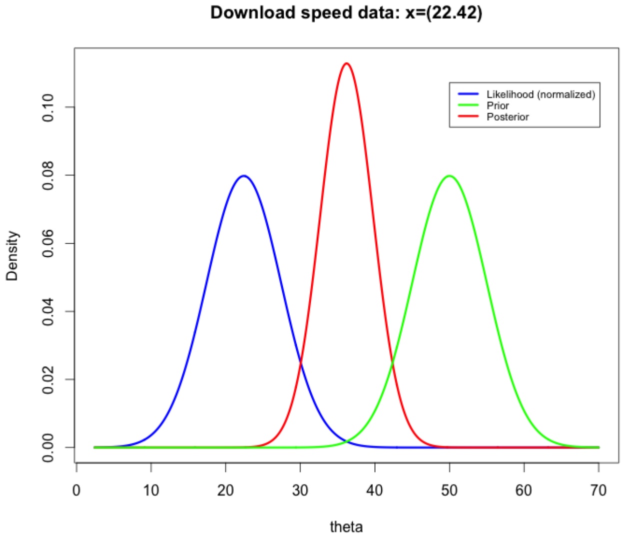
$$w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

- Posterior precision = Data precision + Prior precision.
- Posterior väntevärde $= \frac{\text{Data precision}}{\text{Posterior precision}}$(Data medelvärde). $+ \frac{\text{Prior precision}}{\text{Posterior precision}}$(Prior väntevärde)
- Se sida 344 i Baron för en härledning.

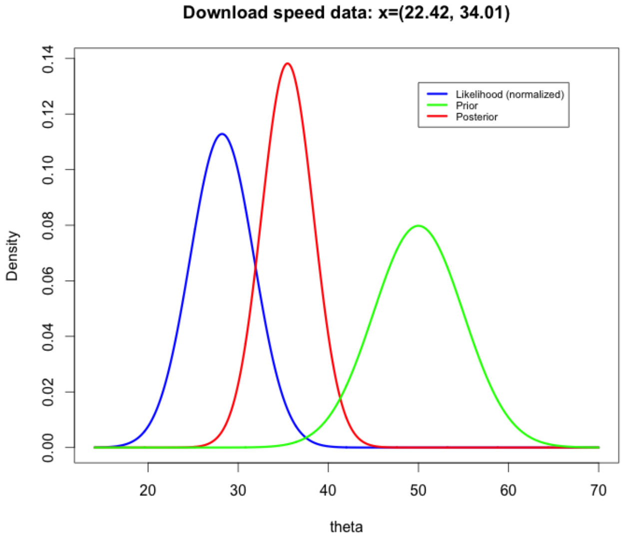
Exempel: Nedladdningshastigheter

- Data: $x = (22.42, 34.01, 35.04, 38.74, 25.15)$.
- Modell: $X_1, \ldots, X_5 \sim N(\theta, \sigma^2)$.
- Antag $\sigma = 5$ (mätningar kan variera $\pm$10MBit med 95% sannolikhet).
- Min prior: $\theta \sim N(50, 5^2)$.

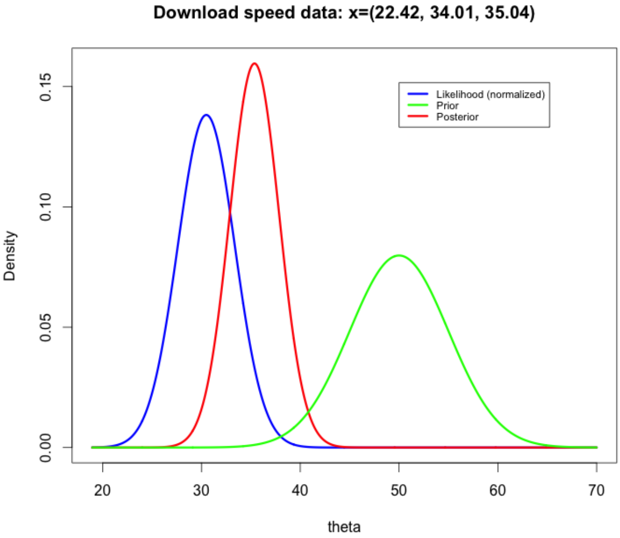
Exempel: Nedladdningshastigheter (n=1)



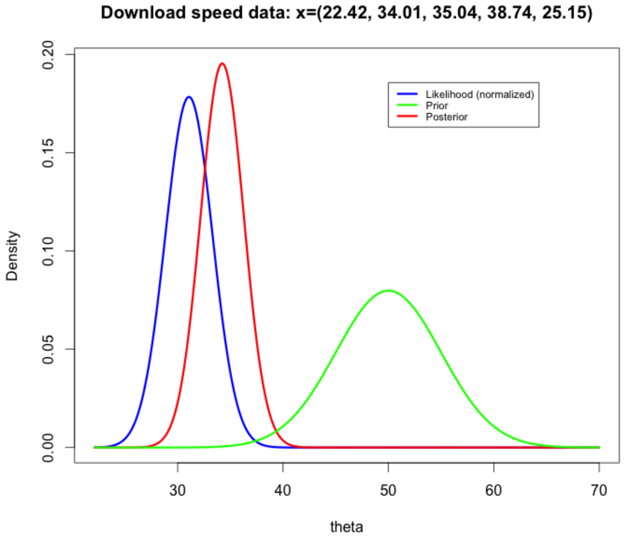
Exempel: Nedladdningshastigheter (n=2)



Exempel: Nedladdningshastigheter (n=3)



Exempel: Nedladdningshastigheter (n=5)



Multinomialmodell med Dirichlet prior

- *Data*: $y = (y_1, \ldots, y_K)$ där y_k = antalet observationer i den k:te klassen.
- Exempel: $K = 8$, och y_k antal som röstar på parti k i en valundersökning med $n = \sum_{k=1}^{K} y_k$ tillfrågade personer.
- Generalisering av binomial till flera klasser.
- **Multinomial modell**:

$$p(y|\theta) \propto \prod_{k=1}^{K} \theta_k^{y_k} \text{ där } \sum_{k=1}^{K} \theta_k = 1$$

- **Konjugerad prior**: *Dirichlet*$(\alpha_1, \ldots, \alpha_K)$

$$p(\theta) \propto \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

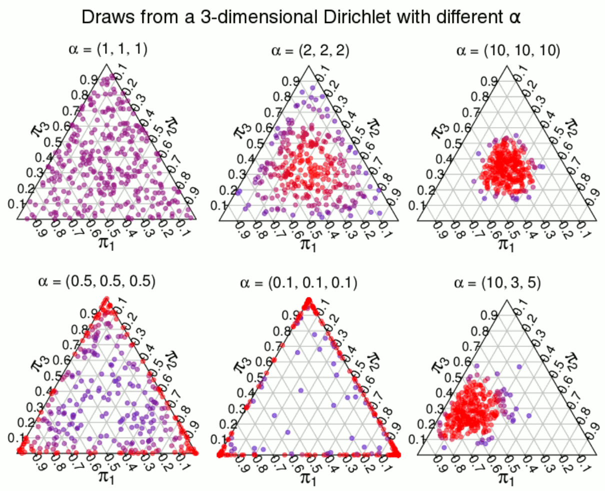
- **Väntevärde** för $\theta = (\theta_1, \ldots, \theta_K) \sim \textit{Dirichlet}(\alpha_1, \ldots, \alpha_K)$

$$\mathbb{E}(\theta_k) = \frac{\alpha_k}{\sum_{j=1}^{K} \alpha_j}$$

- Variansen minskar för större α-värden. **Icke-informativ** prior har små värden, t ex $\alpha_k = 1$ för alla k.

Dirichletfördelningen

- Generalisering av betafördelning till flera dimensioner.



Multinomialmodell med Dirichlet prior

- **Uppdatering** från prior till posterior:
  - Modell: $y = (y_1, \ldots, y_K) \sim \mathit{Multinomial}(n; \theta_1, \ldots, \theta_K)$.
  - Prior: $\theta = (\theta_1, \ldots, \theta_K) \sim \mathit{Dirichlet}(\alpha_1, \ldots, \alpha_K)$.
  - Posterior: $\theta | y \sim \mathit{Dirichlet}(\alpha_1 + y_1, \ldots, \alpha_K + y_K)$.
- **Simulering** från en Dirichlet fördelning:
  - Slumpa $x_1 \sim \mathit{Gamma}(\alpha_1, 1), \ldots, x_K \sim \mathit{Gamma}(\alpha_K, 1)$.
  - Beräkna $z_k = x_k / (\sum_{j=1}^{K} x_j)$.
  - z = $(z_1, \ldots, z_K)$ är $\mathit{Dirichlet}(\alpha_1, \ldots, \alpha_K)$ fördelad.

Exempel: Marknadsandelar

- En undersökning bland 513 smartphone-ägare gav:
  - 180 föredrar en iPhone.
  - 230 föredrar en Androidtelefon.
  - 62 föredrar en Blackberrytelefon.
  - 41 föredrar något annat märke.
- Tidigare undersökning: iPhone 30%, Android 30%, Blackberry 20% och Annat 20%.
- $P(\text{Android har störst marknadsandel} \mid \text{Data})$?
- Min prior: $\alpha_1 = 15, \alpha_2 = 15, \alpha_3 = 10$ och $\alpha_4 = 10$, dvs prior info motsvarar den tidigare undersökningen med 50 svarande.
- Posterior: $(\theta_1, \theta_2, \theta_3, \theta_4)|\mathrm{y} \sim \mathit{Dirichlet}(195, 245, 72, 51)$.

Exempel: Marknadsandelar

```
# Setting up data and prior
y <- c(180,230,62,41) # The cell phone survey data (K=4)
alpha <- c(15,15,10,10) # Dirichlet prior hyperparameters
nIter <- 1000 # Number of posterior draws

# Defining a function that simulates from a Dirichlet distribution
SimDirichlet <- function(nIter, param){
  nCat <- length(param)
  thetaDraws <- as.data.frame(matrix(NA, nIter, nCat)) # Storage.
  for (j in 1:nCat){
    thetaDraws[,j] <- rgamma(nIter,param[j],1)
  }
  for (i in 1:nIter){
    thetaDraws[i,] = thetaDraws[i,]/sum(thetaDraws[i,])
  }
  return(thetaDraws)
}

# Posterior sampling from Dirichlet posterior
thetaDraws <- SimDirichlet(nIter,y + alpha)
# Posterior mean and standard deviation of Androids share (in %)
message(mean(100*thetaDraws[,2]))

## 43.5281011565501

message(sd(100*thetaDraws[,2]))

## 2.12159111747558

# Computing the posterior probability that Android is the largest
PrAndroidLargest <- sum(thetaDraws[,2] > max(thetaDraws[,c(1,3,4)]))/nIter
message(paste('Pr(Android has the largest market share) = ', PrAndroidLargest))

## Pr(Android has the largest market share) =  0.907
```

Bayesiansk estimator, konfidensintervall och hypotestest

- Bayesiansk estimator av θ:

$$E(\theta|x) = \int \theta\pi(\theta|x)d\theta$$

- $[a, b]$ är en 95 % Bayesianskt konfidensintervall om

$$\boldsymbol{P}(\theta \in [a, b]|x) = \int_a^b \pi(\theta|x)d\theta = 0.95$$

- Bayesianskt hypotestest $H_0 : \theta \in \Theta_0$ vs $H_A : \theta \in \Theta_A$. Förkasta H_0 om

$$\pi(\Theta_0|x) < \pi(\Theta_A|x)$$

Översikt

- **Bayesiansk inferens**
- **Bernoullimodell med beta prior**
- **Normalmodell med normal prior**
- **Multinomialmodell med Dirichlet prior**
- **Bayesiansk estimator, konfidensintervall och hypotestest**