

Oulun yliopiston matemaattisten tieteiden laitos/tilastotiede

JOHDATUS BAYESILÄISEEN TILASTOTIETEeseen, kl 2011 (Esa Läärä)

Loppukuulustelu, ma 14.3. klo 14-18 L1

Merkitse vastauspaperiin selvästi, suoritanko kurssin aineopintoihin vai syventäviin opintoihin. VASTAA VIITEEN (5) KYSYMYKSEEN. Tehtävä 1 sisältyy vain aineopintotason tenttiin, kun taas tehtävä 6 vaaditaan vain syventävien opintojen kuulusteluun. Osa muiden tehtävien alakohdista sisältyy vain syventävänä suoritettavaan tenttiin.

Mukana saa olla laskin mutta ei omia muistiinpanoja eikä muutakaan kirjallista materiaalia. Tarvittavat taulukot ovat tehtäväpaperin liitteissä 1-2.

Tausta-aineistoa tehtäviin 1.-3.

Narkolepsiaan sairastuneiden 5-19-vuotiaiden lasten ja nuorten lukumäärät Suomessa vuosijakson 2006-2010 aikana olivat vuosittain seuraavat:

vuosi	2006	2007	2008	2009	2010
lukumäärä	5	6	14	9	62

Lähde: Kansallisen narkolepsiatyöryhmän raportti, 2011 (http://www.thl.fi/fi_FI/web/fi/rokotteet/narkolepsia_ja_sikainfluenssarokote).

Narkolepsiatapausten lukumääristä Suomessa ei ole luotettavia tilastoja ennen vuotta 2006. Ruotsissa diagnosoitiin 2000-luvun puolivälistä alkaen vuosittain 7-16 uutta narkolepsiatapausta tässä ikäluokassa.

Merkitään y_i = ikäluokassa 15-19 vuotta ilmaantuneiden uusien narkolepsiatapausten lukumäärä Suomessa vuonna $i + 2005$; $i = 1, \dots, 5$. Oletetaan, että havainnot y_1, \dots, y_5 ovat toisistaan ehdollisesti riippumattomat siten, että $y_i | \theta_i \sim \text{Poisson}(\theta_i)$, jossa $\theta_i > 0$ jok. $i = 1, \dots, 5$.

TEHTÄVÄT

1. *Vain aineopintoihin:* Tarkastellaan aluksi sellaista yksinkertaista mallia, jossa neljänä ensimmäisenä vuonna 2006-2009 vuotuisilla tapausmäärillä on ollut yhteinen odotusarvo μ eli $\theta_i = \mu$, $i = 1, \dots, 4$. Parametrin μ priorijakaumaksi valitaan Gamma(a_0, b_0), jossa $a_0 > 0$ ja $b_0 > 0$.

- Gamma(a_0, b_0)-jakauma on Poisson(μ)-jakauman konjugaattijakauma. Perustele, miksi näin on.
- Ruotsin kokemusten perusteella μ :n arvoon liittyvä priorinformaatio voidaan Suomen väestöpohjaa varten pukea sellaiseen muotoon, että $\mathbf{E}(\mu) = 6$ ja $\text{var}(\mu) = 6$. Kun μ :n priorijakaumasta oletetaan, että $\mu \sim \text{Gamma}(a_0, b_0)$, niin nämä μ :n odotusarvon ja varianssin arvot vastaavat seuraavia prioriparametrien arvoja: $a_0 = 6, b_0 = 1$. Perustele, miksi näin on.
- Johda annetun havaintoaineiston sekä edellisessä kohdassa määritellyn priorijakauman pohjalta parametrin μ posteriorijakauma, ja laske sen odotusarvo ja keskihajonta.
- Laske normaaliapproksimaatioon perustuen 95% likimääräinen posterioriväli parametrille μ . Vertaa tarkkaan väliin, joka on [5.72, 10.66].

2. Jatkamme narkolepsia-aineiston analyysiä. Seuraavassa on eräs Jags-skripti sekä sen pohjalta tehdyt simulaatioajat ja niiden tuloksia. Kerro tiiviisti mutta riittävän yksityiskohtaisesti, mikä on ollut analyysin kohde, millaisia malli-, priori- ym. oletuksia on tehty, ja mitkä ovat analyysin keskeiset tulokset.

```
write("
  model{
    for( i in 1:length(y) )
      { y[i] ~ dpois(theta[i])
        theta[i] <- beta0 + beta1*(x[i]-mean(x)) }
      beta0 ~ dgamma(6, 1)
      beta1 ~ dnorm(0, 0.05)
      pb1 <- step(beta1)
      tt <- beta0 + beta1*2.5
      yt ~ dpois(tt)
    }
  }, "gampoisl.in.txt")

> library(rjags)
> y <- c(5, 6, 14, 9)
> x <- c(2006, 2007, 2008, 2009)
> gpl <- jags.model("gampoisl.in.txt", data = list(y = y, x = x))
> gpls <- coda.samples(gpl, c("beta0", "beta1", "pb1", "yt"), n.iter = 5000)

> summary(gpls)
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
beta0	8.1001	1.248	0.017651	0.025521
beta1	1.7729	1.148	0.016237	0.019116
pb1	0.9336	0.249	0.003521	0.004836
yt	12.5256	4.839	0.068435	0.072003

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
beta0	5.8093	7.228	8.047	8.927	10.651
beta1	-0.5269	1.012	1.810	2.563	3.935
pb1	0.0000	1.000	1.000	1.000	1.000
yt	4.0000	9.000	12.000	16.000	23.000

3. Oletetaan nyt tehtävän 1 tapaan, että neljänä ensimmäisenä vuonna 2006–2009 vuotuisen tapausmäärien odotusarvo oli samansuuruinen $\mu_0 = \theta_i$, $i = 1, \dots, 4$ mutta vuoden 2010 odotusarvo $\mu_1 = \theta_5$ oli mahdollisesti erilainen. Analysoidaan odotusarvojen μ_1 ja μ_0 suhdetta $\rho = \mu_1/\mu_0$ normaalijakaumaan perustuvalla likimääräisellä menetelmällä. Tässä nojaututaan

siihen, että parametrin $\lambda = \log(\rho)$ suurimman uskottavuuden estimaattorin $\hat{\lambda} = \log(\hat{\rho})$ otanta-jakauma on likimain normaalijakauma odotusarvolla λ ja varianssilla σ^2 , jossa su-estimaattorien ja σ^2 :n lausekkeet ovat

$$\hat{\mu}_0 = \frac{1}{4} \sum_{i=1}^4 y_i, \quad \hat{\mu}_1 = y_5, \quad \hat{\lambda} = \log(\hat{\rho}) = \log(\hat{\mu}_1) - \log(\hat{\mu}_0), \quad \sigma^2 = \frac{1}{4\hat{\mu}_0} + \frac{1}{\hat{\mu}_1},$$

Parametrin λ priorijakaumaksi omaksutaan normaalijakauma prioriodotusarvolla λ_0 ja priorivarianssilla ν_0 .

Näillä oletuksilla parametrin λ posteriorijakauma on likimain normaalijakauma odotusarvolla λ_1 ja varianssilla ν_1 , jossa

$$\nu_1 = \left(\frac{1}{\nu_0} + \frac{1}{\sigma^2} \right)^{-1}, \quad \lambda_1 = \frac{\nu_0^{-1} \lambda_0 + \sigma^{-2} \hat{\lambda}}{\nu_0^{-1} + \sigma^{-2}}.$$

- (a) *Syventäviin opintoihin*: Perustele edellä kerrottu tulos koskien λ :n likimääräistä posteriorijakaumaa.
- (b) Kirjoita Jags-malliskripti, jolla simuloidaan odotusarvojen suhteen $\rho = \mu_1/\mu_0$ posteriorijakaumaa edellä määritellyyn normaaliapproksimaatioon sekä latteaan priorijakaumaan $\lambda \sim N(0, 10^{-6})$ perustuen ja jolla saadaan allaolevat tulokset. Todettakoon, että `prhogt1` tarkoittaa todennäköisyyttä $\mathbb{P}(\rho > 1 \mid y)$. – Tarvittavat havaintovektorin $y = (y_1, \dots, y_5)$ muunnokset kannattaa tehdä tämän skriptin ulkopuolella tavanomaisten R-komentojen avulla.

```
> rhomalli <- jags.model("poisrho.txt", data = list(lamhat = lamhat,
+          SE = SE))
> rhosimu <- coda.samples(rhomalli, c("lambda", "rho", "prhogt1"),
+          n.iter = 10000)

> summary(rhosimu)
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
lambda	1.989	0.1713	0.001713	0.001688
prhogt1	1.000	0.0000	0.000000	0.000000
rho	7.414	1.2781	0.012781	0.012667

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
lambda	1.651	1.875	1.989	2.105	2.324
prhogt1	1.000	1.000	1.000	1.000	1.000
rho	5.211	6.519	7.308	8.210	10.213

- (c) Esitä tiivis sanallinen selostus saaduista tuloksista ja niiden pohjalta tekemistäsi päätelmistä.

4. Olkoon y havaittava satunnaismuuttuja, joka noudattaa binomijakaumaa: $y | \theta \sim \text{Bin}(n; \theta)$, jossa $0 < \theta < 1$.

Klassisessa artikkelissaan Thomas Bayes (1763) johti seuraavan θ :n posteriorijakaumaa koskevan tuloksen

$$\mathbb{P}(u < \theta < v | y; n) = \frac{\binom{n}{y} \int_u^v \theta^y (1 - \theta)^{n-y} d\theta}{\binom{n}{y} \int_0^1 \theta^y (1 - \theta)^{n-y} d\theta}, \quad u, v \in]0, 1[, u < v.$$

- (a) Mihin priorijakaumaan tämä tulos (alkuperäinen Bayesin kaava!) perustuu ja miten se johdetaan annetuista oletuksista ja tunnetuista todennäköisyyslaskennan tuloksista?
- (b) Jorvin sairaalassa Espoossa syntyi naisten päivänä 8.3.2011 yhteensä 14 lasta, joista 13 oli tyttöjä (*Helsingin Sanomat* 10.3.2011). Olkoon $\theta =$ todennäköisyys, että syntyvä lapsi on tyttö. Voidaan osoittaa, että Bayesin käyttämällä priorilla ja em. kaavan nojalla näille havainnoille ehdollistettuna θ :n 95% posterioriväliksi tulee $[0.6805, 0.9834]$. Kommentoi tätä tulosta ja sen pohjana olevia oletuksia. Millaisen priorijakauman itse määrittelisit θ :lle tässä erityisessä sovellustilanteessa?

5. Tee selkoa posteriorijakauman MCMC-simuloinnissa tuotettavien ketjujen suppenemisiagnostiikan periaatteista ja perusmenetelmistä.

6. *Vain syventäviin opintoihin:* Hepatiitti B-virusta (HBV) vastaan kehitetyn rokotteen tehoa tutkittiin Gambiassa. Mitattavana vasteena on vasta-ainemäärä A_{ij} henkilöllä i mittauskertana $j = 1, 2, 3$, jolloin oli kulunut t_{ij} vuorokautta rokotamisesta. Mittauskertoja oli korkeintaan 3 per henkilö. Seuraavassa kuvassa on esitetty kaikkien tutkittujen henkilöiden ($n = 106$) mittaussarjat rokotamisesta kuluneen ajan funktiona.

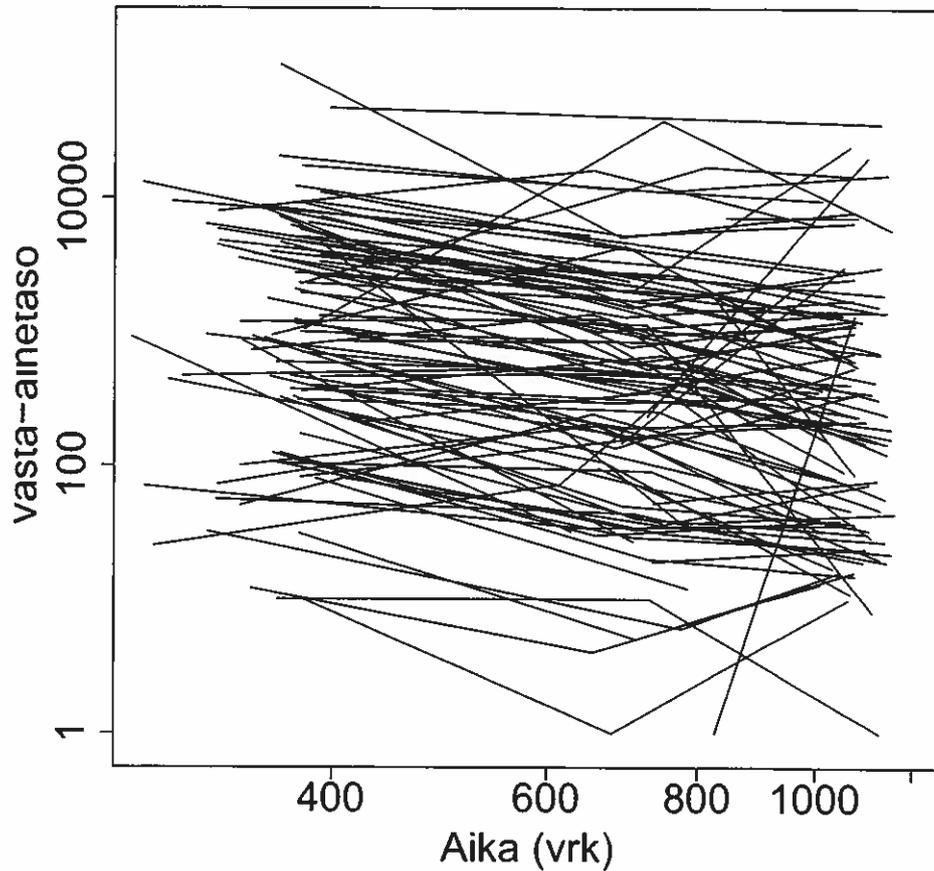
Vasta-ainetasojen logaritmien $y_{ij} = \log(A_{ij})$ malliksi omaksutaan $y_{ij} | \mu_{ij}, \sigma^2 \sim N(\mu_{ij}, \sigma^2)$ ($i = 1, \dots, n; j = 1, 2, 3$) muista havainnoista ehdollisesti riippumatta siten, että odotusarvot riippuvat lineaarisesti ajan logaritmista:

$$\mu_{ij} = \alpha_i + \beta_j \log(t_{ij}/730), \quad i = 1, \dots, n; j = 1, 2, 3.$$

Regressiokertoimien oletetaan noudattavan seuraavia malleja

$$\alpha_i | \mu_\alpha, \sigma_\alpha^2 \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \beta_j | \mu_\beta, \sigma_\beta^2 \sim N(\mu_\beta, \sigma_\beta^2), \quad i = 1, \dots, n,$$

toisistaan ehdollisesti riippumatta. Kaikki tarvittavat hyperpriorijakaumat määritetään latteiksi: odotusarvoille $N(0, 10^{-6})$ ja tarkkuuksille $\text{Gamma}(10^{-3}, 10^{-3})$ toisistaan riippumatta.



Simuloinnin jälkeen saatiin seuraavat tulokset, jossa sigma, sigma.alpha, sigma.beta viittaavat vastaavien varianssien $\sigma^2, \sigma_\alpha^2, \sigma_\beta^2$ neliöjuuriin.

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
mu.alpha	6.0344	0.19187	0.0019187	0.002843
mu.beta	-1.0702	0.14537	0.0014537	0.015156
sigma	1.0029	0.05628	0.0005628	0.001243
sigma.alpha	1.8628	0.14583	0.0014583	0.001799
sigma.beta	0.1971	0.15895	0.0015895	0.017952

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu.alpha	5.66225	5.90517	6.0360	6.1599	6.4103
mu.beta	-1.33997	-1.17220	-1.0801	-0.9688	-0.8002
sigma	0.89972	0.96484	1.0007	1.0382	1.1149
sigma.alpha	1.60403	1.76286	1.8556	1.9545	2.1688
sigma.beta	0.03688	0.08532	0.1452	0.2563	0.6247

- (a) Piirrä suunnattu, sykliton vuokaavio eli DAG, joka havainnollistaa graafisesti näiden oletusten määrittelemän mallikokonaisuuden elementit ja niiden keskinäiset suhteet.
- (b) Kirjoita Jags-skripti, jolla tämän mallin parametrien posteriorijakaumaa voidaan simuloida.
- (c) Tulkitse sanallisesti mallin sisältö sekä saadut tulokset. Kommentoi erityisesti regressio-kertoimille α_i ja β_i tehtyjä oletuksia ja miten niitä voi tässä tilanteessa perustella verrattuna ajateltavissa oleviin vaihtoehtoihin, kuten siihen että $\alpha_i = \alpha$ ja $\beta_i = \beta$, $i = 1, \dots, n$, tai että kaikki α_i :t ja β_i :t olisivat toisistaan täysin erillisiä parametreja.

LIITE 1

Ntzoufras, J. (2009) Bayesian Modeling Using Win Bugs. Wiley, Hoboken NJ.

Table 3.3 Functions available in WinBUGS

WinBUGS Syntax	Function	Description
1. abs(x)	$ x $	Absolute value
2. cloglog(x)	$\log(-\log(1-x))$	Complementary log-log function
3. cos(x)	$\cos(x)$	Cosine function
4. cut(x)		Posterior of x is not updated by the likelihood
5. equals(x1, x2)	$f(x_1, x_2) = 1$ when $x_1 = x_2$ $= 0$ otherwise	Binary indicator function for equal nodes
6. exp(x)	e^x	Exponent value
7. inprod(v1[], v2[])	$\sum_i v_{1i} v_{2i}$	Inner product of two vectors
8. interp.lin(x, v1[], v2[])	$v_{2i} + (v_{2,i+1} - v_{2i})$ $\times (x - v_{1i}) / (v_{1,i+1} - v_{1i})$	Interpolation line
8. inverse(M[,])	A^{-1}	Inverse of a symmetric positive-definite matrix
9. log(x)	$\log(x)$	Logarithm (ln)
10. logdet(M[,])	$\log A $	Logarithm of the determinant of a symmetric positive-definite matrix
11. logfact(k)	$\log(k!)$	Log factorial function of an integer
12. loggam(x)	$\log(\Gamma(x))$	Log gamma function
13. logit(x)	$\log \frac{x}{1-x}$	Logit function
14. max(x1, x2)	$\max(x_1, x_2)$	Maximum of two values
15. mean(v[])	$\bar{v} = \sum_{i=1}^n v_i / n$, where n is the length of vector v	Sample mean
16. min(x1, x2)	$\min(x_1, x_2)$	Minimum of two values
17. phi(x)	$P(X \leq x)$, $X \sim N(0, 1)$	CDF of standardized normal
18. pow(x, z)	x^z	Power function
19. sin(x)	$\sin(x)$	Sine function
20. sqrt(x)	\sqrt{x}	Square root
21. rank(v[], k)	$\sum_i I(v_i \leq v_k)$, where $I(z) = 1$ if z true and 0 otherwise	Rank of s component of a vector
22. ranked(v[], k)	$v_i : \sum_s I(v_s \leq v_i) = k$	Element of a vector with rank s
23. round(x)		Round to the closest integer
24. sd(v[])	$\sqrt{\sum_{i=1}^n (v_i - \bar{v})^2 / (n-1)}$	Sample standard deviation
25. step(x)	$f(x) = 1$ when $x \geq 0$; 0 otherwise	Binary indicator function of positive nodes
26. sum(v[])	$\sum_i v_i$	Sum of a vector's components
27. trunc(x)		Truncation to the closest smaller than x integer

Key: x , z = single real value or logical or mathematical expression; k = single integer value; v = vector; M = matrix.

Table 3.1 Univariate distributions available in WinBUGS

Distribution name	WinBUGS syntax	Probability or density function $f(x)$	Mean	Variance
Discrete distributions				
(1) Bernoulli	x ~ dbern(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$
(2) Binomial	x ~ dbin(p, n)	$n!p^x(1-p)^{n-x}/[x!(n-x)!]$	np	$np(1-p)$
(3) Categorical	x ~ dcat(p[])	p_x	$\sum_{x=1}^K xp_x$	$\sum_{x=1}^K [x - E(x)]^2 p_x$
(4) Negative binomial	x ~ dnegbin(p, r)	$(x+r-1)p^r(1-p)^x/[x!(r-1)!]$	$r(1-p)/p$	$r(1-p)/p^2$
(5) Poisson	x ~ dpois(lambda)	$\exp(-\lambda)\lambda^x/x!$	λ	λ
Continuous distributions				
(6) Beta	x ~ dbeta(a, b)	$\Gamma(a+b)x^{a-1}(1-x)^{b-1}/[\Gamma(a)\Gamma(b)]$	$a/(a+b)$	$ab/[(a+b)^2(a+b+1)]$
(7) Chi-squared	x ~ dchisqr(k)	See gamma($k/2, \frac{1}{2}$)	k	$2k$
(8) Double exponential	x ~ ddexp(mu, tau)	$\frac{1}{2}\tau \exp(-\tau x-\mu)$	μ	$\sqrt{2}/\tau$
(9) Exponential	x ~ dexp(lambda)	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
(10) Gamma	x ~ dgamma(a, b)	$b^a x^{a-1} e^{-bx} / \Gamma(a)$	a/b	a/b^2
(11) Generalized gamma	x ~ gen.gamma(a, b, r)	$\tau b(bx)^{r-1} \exp[-(bx)^r] / \Gamma(a)$	$\Gamma(a+1/r)/[b\Gamma(a)]$	$[\Gamma(a+2/r^2)\Gamma(a) - \Gamma(a+1/r)^2]/[\lambda\Gamma(a)]^2$
(12) Log-normal	x ~ dlnorm(mu, tau)	$\sqrt{\tau/(2\pi)} x^{-1} \exp[-\tau/2(\log x - \mu)^2]$	$e^{\mu+1/(2\tau)}$	$(e^{1/\tau} - 1)e^{2\mu+1/\tau}$
(13) Logistic	x ~ dlogis(mu, tau)	$\tau e^{\tau(x-\mu)} [1 + e^{\tau(x-\mu)}]^{-2}$	μ	$\pi^2/[3\tau^2]$
(14) Normal	x ~ dnorm(mu, tau)	$\sqrt{\tau/(2\pi)} \exp[-\tau(x-\mu)^2/2]$	μ	$1/\tau$
(15) Pareto	x ~ dpar(a, c)	$ac^a x^{-a-1}$	$ab/(a-1)$	$ab^2/[(a-1)^2(a-2)]$
(16) Student's t	x ~ dt(mu, tau, v)	$\Gamma[(v+1)/2] \sqrt{\tau/(2\pi)} [\Gamma(v/2)]^{-1} \times [1 + \tau v^{-1}(x-\mu)^2]^{-(v+1)/2}$	μ	$v\tau^{-1}/(v-2)$
(17) Uniform	x ~ dunif(a, b)	$1/(b-a)$	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$
(18) Weibull	x ~ dweib(v, lambda)	$v\lambda x^{v-1} \exp(-\lambda x^v)$	$\lambda^{-1/v} \Gamma(1+v^{-1})$	$[\Gamma(1+2v^{-1}) - \Gamma(1+v^{-1})^2] \lambda^{-2/v}$

Key: $p \in (0, 1)$; (1) $x = 0, 1$; (2) $x = 0, 1, \dots, n$, $p \in (0, 1)$; (3) $p[]$ is a vector of dimension K , $x = 1, 2, \dots, K$ and elements $p[x] = p_x \in (0, 1)$, $\sum_{x=1}^K p_x = 1$; (4) $x = 0, 1, 2, \dots$; $r = 1, 2, \dots$; (5) $x = 0, 1, 2, \dots$; $\lambda > 0$; (6) $x \in (0, 1)$, $a, b > 0$; (8) $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\tau > 0$; (9) $x > 0$, $\lambda > 0$; (10, 11) $x > 0$, $a, b, r > 0$; (12) $x > 0$, $\mu \in \mathbb{R}$, $\tau > 0$; (13, 14, 16) $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\tau > 0$; (15) $x > c$, $a, c > 0$; (17) $x \in (a, b)$, $a, b \in \mathbb{R}$, $a < b$; (18) $x > 0$, $v, \lambda > 0$.

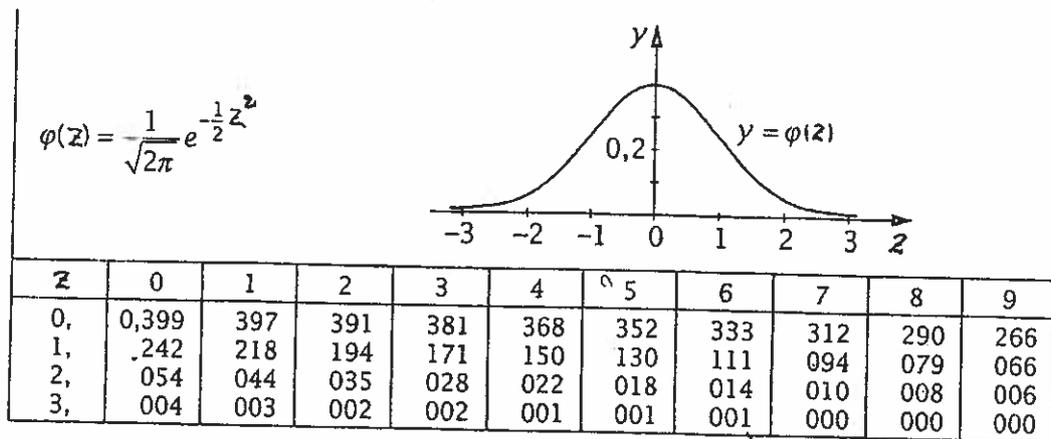
τv^{-1}
 $\frac{1}{2}(a+b)$

Table 3.2 Multivariate distributions available in WinBUGS

Distribution name	WinBUGS syntax	Probability or density function $f(x)$	Mean	Variance/covariance
(19) Multinomial	$x[1:K] \sim \text{dmulti}(p[], N)$	$N! \left(\prod_{i=1}^K x_i! \right)^{-1} \prod_{i=1}^K p_i^{x_i}$	$E(X_i) = Np_i$	$V(X_i) = Np_i(1 - p_i)$ $\text{Cov}(X_i, X_j) = -Np_i p_j$
Discrete distributions				
(20) Dirichlet	$x[1:K] \sim \text{ddirch}(a[])$	$\Gamma(a) \left[\prod_{i=1}^K \Gamma(a_i) \right]^{-1} \prod_{i=1}^K x_i^{a_i - 1}$	$E(X_i) = a_i/a$	$V(X_i) = a_i(a - a_i)/[a^2(a + 1)]$ $\text{Cov}(X_i, X_j) = -a_i a_j/[a^2(a + 1)]$
(21) Multivariate normal	$x[1:K] \sim \text{dmnorm}(\mu[], T[,])$	$(2\pi)^{-K/2} T ^{1/2} \exp \left[-\frac{1}{2} (x - \mu)^T T (x - \mu) \right]$	$E(X) = \mu$	$V(X) = T^{-1}$
(22) Multivariate Student's t	$x[1:K] \sim \text{dmt}(\mu[], T[,], v)$	$\left\{ (v\pi)^{-K/2} \Gamma[(v+K)/2] / \Gamma(v/2) \right\}$ $\times \left[1 + v^{-1} (x - \mu)^T T (x - \mu) \right]^{-(v+K)/2}$	$E(X) = \mu$	$V(X) = v(v - 2)^{-1} T^{-1}$
(23) Wishart	$x[1:K, 1:K] \sim \text{dwish}(R[,], v)$	$ R ^{v/2} x ^{(v-K-1)/2} \exp \left[-\frac{1}{2} \text{Tr}(Rx) \right]$	$E(X_{ij}) = vA_{ij}$	$\text{Cov}(X_{ij}, X_{km}) = v(A_{ik}A_{jm} + A_{im}A_{jk})$

Key: (19) $x[]$ and $p[]$ are vectors of dimension K with elements $x[i] = x_i \in (0, 1, 2, \dots)$ and $p[i] = p_i \in (0, 1)$ with $\sum_{i=1}^K p_i = 1$; (20) $x[]$ and $a[]$ are vectors of dimension K with elements $x[i] = x_i \in (0, 1)$ and $a[i] = a_i > 0$ with $\sum_{i=1}^K a_i = 1$ and $a = \sum_{i=1}^K a_i$; (21, 22) $x[]$ and $\mu u[]$ (x and μ) are vectors of dimension K with elements $x[i] = x_i \in (0, 1)$ and $\mu[i] = \mu_i \in \mathbf{R}$. $T[]$ (and T) is a $K \times K$ symmetric precision matrix with elements $T[i, j] = \tau_{ij} > 0$; $v > 0$; (23) $x[,]$ and $R[,]$ are $K \times K$ positive-definite (symmetric) matrices with elements $x[i, j] = x_{ij}$ and $R[i, j] = R_{ij}$; A_{ij} are the element of the matrix $A = R^{-1}$; $v > 0$. Also μ_i : mean of X_i ; $V(X_i)$, $\text{Cov}(X_i, X_j)$: variance of X_i ; $\text{Cov}(X_i, X_j)$: covariance between X_i and X_j .

LIITE 2: Normaalijakauman tiheysfunktio



Normaalijakauman kertymäfunktio

