STK4011 and STK9011
Autumn 2016

Univariate distributions

Covers most of sections 3.2 and 3.3.
(and parts of sections 1.5,1.6, 2.2 and 2.3)

Ørnulf Borgan
Department of Mathematics
University of Oslo

Discrete distributions

A random variable X is discrete if its range $\mathcal{X} = \{x_1, x_2, ...\}$ is countable (finite or countably infinite)

Cumulative distribution function (cdf)

$$F_X(x) = P(X \leq x)$$

is a step function

Probability mass function (pmf)

$$f_X(x) = P(X = x) = F_X(x) - F_X(x-) \text{ for } x \in \mathcal{X}$$

Note that $f_X(x) = 0 \text{ for } x \notin \mathcal{X}$

The expected value or mean of $g(X)$ is given by

$$\mathrm{E}g(X) = \sum_{x \in \mathcal{X}} g(x) f_X(x)$$

provided that $\sum_{x \in \mathcal{X}} |g(x)| f_X(x) < \infty$

In particular

$$\mu = \mathrm{E}X = \sum_{x \in \mathcal{X}} x\, f_X(x)$$

$$\sigma^2 = \mathrm{Var}\, X = \mathrm{E}\{(X - \mu)^2\} = \mathrm{E}(X^2) - \mu^2$$

Moment generating function (mgf)

$$M_X(t) = \mathrm{E}\, e^{tX} = \sum_{x \in \mathcal{X}} e^{tx} f_X(x)$$

assuming that the expected value exists for all *t* in an open interval that contains zero

Note that $\mathrm{E}\, X^n = M_X^{(n)}(0)$

Hypergeometric distribution

We have an urn with N balls

M balls are red and $N-M$ balls are green

We select at random K balls

X is the number of red balls we select

Probability mass function (pmf)

$$f_X(x) = \frac{\binom{M}{x} \cdot \binom{N-M}{K-x}}{\binom{N}{K}}$$

$$x = 0, 1, \ldots., K$$

$$M - (N-K) \le x \le M$$

$$\mathrm{E}X = K\frac{M}{N} \qquad \mathrm{Var}\,X = K\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-K}{N-1}$$

Binomial distribution

We have n independent and identical Bernoulli trials with success probability p

X is the number of successes in the n trials

Probability mass function (pmf)

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

We have:

$$\mathrm{E}X = np \qquad \text{(example 2.2.3)}$$

$$\mathrm{Var}\, X = np(1-p) \qquad \text{(example 2.3.5)}$$

$$M_X(t) = \left[pe^t + (1-p) \right]^n \qquad \text{(example 2.3.9)}$$

Poisson distribution

X is Poisson distributed with parameter λ if its pmf takes the form

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0, 1, \ldots.$$

We have:

$$\mathrm{E}X = \mathrm{Var}\, X = \lambda$$

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Relation with binomial distribution:

$$\binom{n}{x} p^x (1-p)^{n-x} \to \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{when } n \to \infty,\ np \to \lambda$$

Negative binomial distribution

We have a sequence of independent Bernoulli trials with success probability p

X is the trial at which the r-th success occurs

Probability mass function (pmf)

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad x = r, r+1, \ldots.$$

Alternative formulation: $Y = X - r$ has pmf

$$f_Y(y) = \binom{r+y-1}{y} p^r (1-p)^y \qquad y = 0, 1, \ldots.$$

$$\mathrm{E}Y = \frac{r(1-p)}{p} = \mu \qquad \mathrm{Var}\,Y = \frac{r(1-p)}{p^2} = \mu + \frac{1}{r}\mu^2$$

Geometric distribution

We have a sequence of independent Bernoulli trials with success probability p

X is the trial at which the first success occurs

Probability mass function (pmf)

$$f_X(x) = p(1-p)^{x-1} \qquad x = 1, 2, \ldots.$$

$$\mathrm{E}X = \frac{1}{p}$$

$$\mathrm{Var}\, X = \frac{1-p}{p^2}$$

Alternatively we may consider $Y = X - 1$

Continuous distributions

If the cdf $F_X(x) = P(X \le x)$ is continuous, then X is a continuous random variable

For a continuous random variable (strictly speaking absolutely continuous random variable) we have

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

where $f_X(x)$ is the probability density function (pdf)

Further $f_X(x) = F'_X(x)$

The expected value or mean of $g(X)$ is given by

$$\mathrm{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

provided that $\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\,dx < \infty$

In particular

$$\mu = \mathrm{E}X = \int_{-\infty}^{\infty} x\, f_X(x)\,dx$$

$$\sigma^2 = \mathrm{Var}\, X = \mathrm{E}\{(X-\mu)^2\} = \mathrm{E}(X^2) - \mu^2$$

Note that if $\int_{-\infty}^{\infty} |x|^k\, f_X(x)\,dx < \infty$ for a $k > 0$, then

$\int_{-\infty}^{\infty} |x|^m\, f_X(x)\,dx < \infty$ for all m with $0 < m < k$

Thus if $\mathrm{E}(X^k)$ exists, then $\mathrm{E}(X^m)$ exists for $0 < m < k$

Moment generating function (mgf)

$$M_X(t) = \mathrm{E}\, e^{tX} = \int_{-\infty}^{\infty} e^{tx}\, f_X(x)\, dx$$

assuming that the expected value exists for all *t* in an open interval that contains zero

Note that $\mathrm{E}\, X^n = M_X^{(n)}(0)$

Normal distribution

If the pdf takes the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

then X is normally distributed with mean μ and variance σ^2

Short $X \sim n(\mu, \sigma^2)$

We have that

$$\mathrm{E}X = \mu$$

$$\mathrm{Var}\, X = \sigma^2$$

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

The variable

$$Z = \frac{X - \mu}{\sigma}$$

is standard normally distributed: $Z \sim n(0,1)$

To prove that the normal pdf integrates to 1, we have to prove that

$$\int_0^\infty e^{-z^2/2}\, dz = \sqrt{\frac{\pi}{2}}$$

or equivalently that

$$\left(\int_0^\infty e^{-z^2/2}\, dz \right)^2 = \frac{\pi}{2}$$

The variable

$$Z = \frac{X - \mu}{\sigma}$$

is standard normally distributed: $Z \sim n(0,1)$

To prove that the normal pdf integrates to 1, we have to prove that

$$\int_0^\infty e^{-z^2/2}\, dz = \sqrt{\frac{\pi}{2}}$$

or equivalently that

$$\left(\int_0^\infty e^{-z^2/2}\, dz \right)^2 = \frac{\pi}{2}$$

Laplace distribution (double exponential distribution)

The standard Laplace distribution is given by the pdf

$$f(z) = \frac{1}{2} e^{-|z|} \qquad \text{for} \quad -\infty < z < \infty$$

If Z has this pdf, then $\mathrm{E}Z = 0$ and $\mathrm{Var}\, Z = 2$

Now $X = \sigma Z + \mu$ has pdf

$$f_X(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} \qquad \text{for} \quad -\infty < x < \infty$$

Note that $\mathrm{E}X = \mu$ and $\mathrm{Var}\, X = 2\sigma^2$

Cauchy distribution

The standard Cauchy distribution is given by the pdf

$$f(z) = \frac{1}{\pi} \frac{1}{1+z^2} \qquad \text{for} \quad -\infty < z < \infty$$

One may show that $\mathrm{E}|Z| = \infty$ (example 2.2.4), so for the Cauchy distribution the mean does not exist

Now $X = Z + \theta$ has pdf

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2} \qquad \text{for} \quad -\infty < x < \infty$$

The mean and variance of X do not exist

Exponential distribution

The standard exponential pdf takes the form

$$f(z) = \begin{cases} e^{-z} & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$$

If Z has this pdf, then $\mathrm{E}Z = 1$ and $\mathrm{Var}\, Z = 1$

Then $X = \beta Z$ is exponentially distributed with scale parameter β :

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta} \qquad \text{for} \quad x > 0$$

We have $\mathrm{E}X = \beta$ and $\mathrm{Var}\, X = \beta^2$

Gamma distribution

The standard gamma pdf with shape parameter $\alpha > 0$ is given by

$$f(z) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-z} & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here the gamma function $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz$ has the following properties:

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

$$\Gamma(n+1) = n! \quad \text{when } n \text{ is an integer}$$

$$\Gamma(1) = 1$$

$$\Gamma(1/2) = \sqrt{\pi}$$

Then $X = \beta Z$ is gamma distributed with shape parameter α and scale parameter β :

$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{for } x > 0$$

Short $X \sim gamma(\alpha, \beta)$

We have

$$\mathrm{E}X = \alpha\beta \qquad \mathrm{Var}\, X = \alpha\beta^2$$

$$M_X(t) = \left(\frac{1}{1-\beta t} \right)^\alpha \quad \text{for } t < \frac{1}{\beta}$$

If $X \sim gamma(p/2, 2)$ we say that X is chi squared distributed with p degrees of freedom (p integer)

Lognormal distribution

If $\log X \sim n(\mu, \sigma^2)$, then X has a lognormal distribution

The pdf takes the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \frac{1}{x} e^{-(\log x - \mu)^2/(2\sigma^2)} \qquad \text{for } x > 0$$

We have

$$\mathrm{E}X = e^{\mu + \sigma^2/2}$$

$$\mathrm{Var}\, X = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$$

Uniform distribution

If the pdf takes the form

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

then X is uniformly distributed over $[a , b]$

We have:

$$\mathrm{E}X = \frac{b+a}{2}$$

$$\mathrm{Var}\, X = \frac{(b-a)^2}{12}$$

Beta distribution

The beta distribution is a generalization of the uniform distribution on $[0\,,\,1]$

The $beta(\alpha,\beta)$ - distribution has pdf

$$f_X(x) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } \ 0 < x < 1$$

Here the beta function is given by

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

We have (cf. exercise to week 36):

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

If $X \sim beta(\alpha, \beta)$ we have that

$$\mathrm{E}X^n = \frac{B(\alpha+n, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$$

From this we obtain

$$\mathrm{E}X = \frac{\alpha}{\alpha+\beta}$$

$$\mathrm{Var}\,X = \frac{\alpha\beta}{\alpha+\beta\ ^{2}\ \alpha+\beta+1}$$

Exponential families

Covers section 3.4

Many pdfs or pmfs may be expressed on the form

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) \qquad (3.4.1)$$

where $h(x) \geq 0$ and $t_1(x),....,t_k(x)$ are real-valued functions of x and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}),...,w_k(\boldsymbol{\theta})$ are real valued functions of the possibly vector-valued parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_d), \quad d \leq k$

We say that (3.4.1) defines an exponential family of distributions

Example: normal distribution

Consider the normal pdf

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

This may be written

$$f(x \mid \mu, \sigma^2) = \underbrace{1}_{h(x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2} \right)}_{c(\mu,\sigma)} \exp\left\{ \underbrace{\frac{1}{\sigma^2}}_{w_1(\mu,\sigma)} \underbrace{\left(-\frac{x^2}{2} \right)}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\mu,\sigma)} \underbrace{x}_{t_2(x)} \right\}$$

which is of the form (3.4.1)

Example: exponential distribution

Consider the exponential pdf

$$f(x|\beta)=\begin{cases}\dfrac{1}{\beta}\exp\left(-\dfrac{x}{\beta}\right) & \text{if } x>0\\ 0 & \text{otherwise}\end{cases}$$

For all x this may be written

$$f(x|\beta)=\underbrace{I_{\{x>0\}}(x)}_{h(x)}\underbrace{\frac{1}{\beta}}_{c(\beta)}\exp\left(\underbrace{-\frac{1}{\beta}}_{w_1(\beta)}\underbrace{x}_{t_1(x)}\right)$$

which is of the form (3.4.1)

Example: binomial distribution

Consider the binomial pmf

$$f(x \mid p) = \binom{n}{x} p^{x}(1-p)^{n-x} \qquad x = 0,1,\ldots,n$$

For all x this may be written

$$f(x \mid p) = \underbrace{I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}}_{h(x)} \underbrace{(1-p)^{n}}_{c(p)} \exp\left\{ \underbrace{\log\left(\frac{p}{1-p}\right)}_{w_1(p)} \underbrace{x}_{t_1(x)} \right\}$$

which is of the form (3.4.1)

In chapter 7 we will prove the following result (which may also be known from earlier courses):

If X is a random variable with pdf or pmf $f(x \mid \boldsymbol{\theta})$, and if $\int f(x \mid \boldsymbol{\theta})\, dx$ or $\sum f(x \mid \boldsymbol{\theta})$ may be differentiated twice with respect to θ_j by changing the order of integration/summation and differentiation, then we have

$$\mathrm{E}\left(\frac{\partial}{\partial \theta_j} \log f(X \mid \boldsymbol{\theta}) \right) = 0$$

$$\mathrm{Var}\left(\frac{\partial}{\partial \theta_j} \log f(X \mid \boldsymbol{\theta}) \right) = -\mathrm{E}\left(\frac{\partial^2}{\partial \theta_j^2} \log f(X \mid \boldsymbol{\theta}) \right)$$

If we use this result for the exponential family of distributions, we obtain:

Theorem 3.4.2
If X is a random variable with pdf or pmf of the form (3.4.1), then

$$\mathrm{E}\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(X)\right)=-\frac{\partial}{\partial\theta_j}\log c(\boldsymbol{\theta})$$

$$\mathrm{Var}\left(\sum_{i=1}^{k}\frac{\partial w_i(\boldsymbol{\theta})}{\partial\theta_j}t_i(X)\right)=-\frac{\partial^2}{\partial\theta_j^2}\log c(\boldsymbol{\theta})-\mathrm{E}\left(\sum_{i=1}^{k}\frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial\theta_j^2}t_i(X)\right)$$

Example: exponential distribution

We have the pdf

$$f(x\,|\,\beta) = I_{\{x>0\}}(x)\,\frac{1}{\beta}\exp\left(-\frac{1}{\beta}x\right)$$

We have $c(\beta) = \frac{1}{\beta}, \quad w_1(\beta) = -\frac{1}{\beta}, \quad t_1(x) = x$

By Theorem 3.4.2 it follows that

$$\mathrm{E}\left(\frac{1}{\beta^2}X\right) = \frac{1}{\beta} \qquad \mathrm{Var}\left(\frac{1}{\beta^2}X\right) = -\frac{1}{\beta^2} - \mathrm{E}\left(-\frac{2}{\beta^3}X\right)$$

This gives

$$\mathrm{E}X = \beta \qquad \mathrm{Var}\,X = \beta^2$$

Example: binomial distribution

We have the pmf

$$f(x \mid p) = I_{\{0,1,\ldots,n\}}(x) \binom{n}{x} (1-p)^n \exp\left\{ \log\left(\frac{p}{1-p} \right) x \right\}$$

Note that

$$c(p) = (1-p)^n, \quad w_1(p) = \log\left(\frac{p}{1-p} \right), \quad t_1(x) = x$$

$$\frac{d}{dp} \log c(p) = \frac{-n}{1-p}, \qquad \frac{d}{dp} w_1(p) = \frac{1}{p(1-p)}$$

$$\frac{d^2}{dp^2} \log c(p) = \frac{-n}{(1-p)^2}, \qquad \frac{d^2}{dp^2} w_1(p) = \frac{-(1-2p)}{p^2(1-p)^2}$$

By Theorem 3.4.2 we have

$$\mathrm{E}\left(\frac{1}{p(1-p)}X\right)=\frac{n}{1-p}$$

$$\mathrm{Var}\left(\frac{1}{p(1-p)}X\right)=\frac{n}{1-p\ {}^{2}}-\mathrm{E}\left(\frac{-(1-2p)}{p^{2}(1-p)^{2}}X\right)$$

This gives

$$\mathrm{E}X=np$$

$$\mathrm{Var}\,X=np(1-p)$$

So far we have used the parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$

The parameter space for (3.4.1) is (usually) given as

$$\Theta = \left\{ \boldsymbol{\theta} : \int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x) \right) dx < \infty \right\} = \boldsymbol{\theta} : c(\boldsymbol{\theta}) > 0$$

(in the discrete case the integral is replaced by a sum)

Sometimes an exponential family is reparametrized as

$$f(x \mid \boldsymbol{\eta}) = h(x) c^*(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x) \right)$$

Here the $h(x)$ and $t_i(x)$ functions are as before, and

$\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ is the natural parameter

The natural parameter space is given as

$$\mathcal{H} = \left\{ \boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k) : \int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i \, t_i(x) \right) dx < \infty \right\}$$

Note that

$$c^*(\boldsymbol{\eta}) = \left[\int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i \, t_i(x) \right) dx \right]^{-1}$$

Example: binomial distribution

We have the pmf

$$f(x\,|\,p) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}(1-p)^n \exp\left\{\log\left(\frac{p}{1-p}\right)x\right\}$$

$$= I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}c(p)\exp\left\{w_1(p)\,x\right\}$$

where $c(p) = (1-p)^n$ and $w_1(p) = \log\left(p/(1-p)\right)$

The original parameter space is $\{p : 0 < p < 1\} = (0,1)$

The natural parameter is $\eta = \log\left(p/(1-p)\right) \in (-\infty, \infty)$

Note that $p = e^{\eta}/(1+e^{\eta})$ and that the pmf becomes

$$f(x\,|\,\eta) = I_{\{0,1,\ldots,n\}}(x)\binom{n}{x}\left(\frac{1}{1+e^{\eta}}\right)^n \exp\left\{\eta x\right\}$$

Example: normal distribution

The normal pdf may be written

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

Usually the original parameter space is given by

$$(\mu,\sigma^2):\ -\infty<\mu<\infty,\ \sigma^2>0$$

The natural parameters are $\eta_1=1/\sigma^2$ and $\eta_2=\mu/\sigma^2$

The natural parameter space is

$$(\eta_1,\eta_2):\ \eta_1>0,\ -\infty<\eta_2<\infty$$

The pdf may be written

$$f(x|\eta_1,\eta_2)=\frac{\sqrt{\eta_1}}{\sqrt{2\pi}}\exp\left(-\frac{\eta_2^2}{2\eta_1}\right)\exp\left(\eta_1\left(-\frac{x^2}{2}\right)+\eta_2 x\right)$$

In (3.4.1) it is usually the case that the dimension of the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_d)$ is equal to k, i.e. $d = k$

Then we have a full exponential family

If $d < k$, we have a curved exponential family

In the example on the previous slide, we have a full exponential family with parameter space

$$(\mu, \sigma^2): -\infty < \mu < \infty,\ \sigma^2 > 0$$

If we consider a normal model with $\sigma^2 = k\mu^2$, for a known constant k, we get a curved exponential family with parameter space $(\mu, k\mu^2): -\infty < \mu < \infty$

Location and scale families

Note that if

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

is the standard normal density, then the $n(\mu, \sigma^2)$ density may be written

$$f_X(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

This is an example of a location and scale family of distributions

In general we may obtain a location and scale family of distributions by starting with a standard pdf $f(z)$

The location and scale family is then given by all pdfs of the form

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

(one may easily check that these are pdfs)

μ is the location parameter $(-\infty < \mu < \infty)$

σ is the scale parameter $(\sigma > 0)$

The following result relates the location and scale family to the transformation of random variables (just as for the normal case)

Theorem 3.5.6

Let $f(\cdot)$ be any pdf. Let μ be any real number and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$

Note that if $\mathrm{E}Z$ and $\mathrm{Var}\,Z$ exist, then $\mathrm{E}X = \sigma \mathrm{E}Z + \mu$ and $\mathrm{Var}\,X = \sigma^2 \mathrm{Var}\,Z$