AMS 276
Lecture 2: Parametric Models

Fall 2016

♣ Common Parametric Models for Survival Data:

- Consider univariate right censored survival data.
- Recall that we have
  - ⋆⋆ t_i: a continuous nonngeative random variable representing survival time of subject i
  - ⋆⋆ c_i: a fixed censoring time
  - ⋆⋆ Data includes $y_i = \min(t_i, c_i)$ and a binary indicator ν_i (1 for not censored, 0 for censored)
- We are in KM Chapter 2 & ICS Chapter 2 (but skip the regression part).

• Nonparametric, semiparametric and parametric models can be used as a model for T.

• We will look at some common parametric models.

e.g.: exponential, Weibull, gamma, log normal, log logistic...

• See Table 2.2 of Klein and Moeschberger (p38)

⋆⋆ Hazard rate, Survival function, pdf are listed for common parametric models.

♣ Bayesian Paradigm (ICS 1.5–1.8):

- Data: $\mathcal{D} = (n, \mathbf{y}, \boldsymbol{\nu})$ & $\boldsymbol{\theta}$: a vector of indexing parameters.
- Likelihood function: $\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D})$
- Prior distribution: $\pi(\boldsymbol{\theta})$
- Posterior distribution:

$$\pi(\boldsymbol{\theta} \mid \mathcal{D}) \propto \underbrace{\pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D})}_{P(D)}$$

⋆⋆ We can sample from $\pi(\boldsymbol{\theta} \mid \mathcal{D})$ without knowing the normalizing constant through MCMC.

- Predictive distribution:

$$\pi(t^{\text{new}} \mid \mathcal{D}) = \int f(t^{\text{new}} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathcal{D})d\boldsymbol{\theta}.$$

♣ Informative Prior Elicitation

• Suppose we have data from previous similar studies (historical data). Recall our previous example, Melanoma data. E 1684 E 1690

• Let $\mathcal{D}_0 = (n_0, \mathbf{y}_0, \boldsymbol{\nu}_0)$ denote historical data.

• $\mathcal{D}_0$ can be very helpful in interpreting the results of the current study.

♣ Why not just noninformative (proper or improper) priors?

• Cannot be used in all applications, such as model selection and model comparison.

• Bayes factors are sensitive to the choices of hyperparamerers of noninformative prior priors.

• Convergence problems for the Gibbs sampler

• Do not make use of real prior information.

" Power Prior distributions for Regression models " by Ibrahim & Chen (2000)

- One way to derive an informative prior is to consider the **power prior**;

$$\underbrace{\pi(\boldsymbol{\theta} \mid \mathcal{D}_0)} \propto \underbrace{\pi_0(\boldsymbol{\theta} \mid c_0)}_{\text{initial prior for } \theta}\{\overbrace{\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D}_0)}^{\text{likelihood of historical data}}\}^{a_0}$$

★★ c_0: a specified hyperparameter for the initial prior.

★★ $0 \leq a_0 \leq 1$: a scalar prior parameter that weights the historical data relative to the likelihood of the current study.

- $a_0 = 0$: no incorporation of historical data
- $a_0 = 1$: the prior for the current study becomes the posterior distribution from historical data

★★ We can consider a prior for a_0, $\pi(a_0 \mid \gamma_0)$ where γ_0 is a specified hyperparameter (like Be)

$a_0 = 0 \Rightarrow \pi(\theta \mid D_0) \propto \underbrace{\pi_0(\theta \mid c_0)} \times 1$: no use of historical data

$a_0 = 1 \Rightarrow \pi(\theta \mid D_0) \propto \pi_0(\theta \mid c_0) \times \mathcal{L}(\theta \mid D_0)$: full use of historical data

$$\pi(\theta \mid D_0, D) \propto \underbrace{\pi_0(\theta \mid c_0)\{\mathcal{L}(\theta \mid D_0)\}^{a_0}} \times \mathcal{L}(\theta \mid D)$$

$$\pi(a_0 \mid -) = p_0 I(a_0 = 0) + p_1 I(a_0 = 1) + (1 - p_0 - p_1)\, \underline{Be}$$

♣ Why Bayes?

• Censoring $\Rightarrow$ complex data structure $\Rightarrow$ But, straightforward to take care of this in the Bayesian framework (NO asymptotic calculation!)

• Incorporate historical data in a natural way (We just saw one way of doing this).

• Flexibility: model comparison, missing covariate and response data.....

♣ What is coming next?

- Will look at common parametric models for likelihood; Exponential, Weibull, Extreme Value, Log-Normal, Gamma.

- For each, we will consider a Bayesian way to extend the model.

♣ **Exponential** distribution:

$$y_i \mid \lambda \overset{iid}{\sim} \mathrm{E}(\lambda), \quad i = 1, \ldots, n \;\text{ with } \lambda > 0.$$

- Density function:

$$f(y \mid \lambda) = \begin{cases} \lambda \exp(-\lambda y) & \text{if } y > 0, \\ 0, & \text{o.w.} \end{cases}$$

- Mean $1/\lambda$ and variance $1/\lambda^2$. $= 1 - F(y \mid \lambda) = 1 - \int_0^y \lambda e^{-\lambda u}\, du$
- Survival function $S(y \mid \lambda) = \exp(-\lambda y)$ for $y > 0$. $= 1 - (-e^{-\lambda u})\big|_0^y$ $= e^{-\lambda y},\ y > 0$
- Hazard function $h(y \mid \lambda) = \lambda$: Constant hazard!
- Special case of both the Weibull and gamma distributions (will see later).

$h(y \mid \lambda) = \dfrac{f(y \mid \lambda)}{S(y \mid \lambda)} = \dfrac{\lambda \cancel{e^{-\lambda y}}}{\cancel{e^{-\lambda y}}} = \lambda$

$$h(y|\lambda) = \lim_{dy \to 0^+} \frac{P(Y \in [y, y+dy) \mid Y \geq y)}{dy}$$

$$= \lim_{dy \to 0^+} \frac{P(Y \in [y, y+dy))}{dy}$$

♣ **Exponential** distribution (contd)

⋆⋆ Hazard function $h(y \mid \lambda) = \lambda$: Constant hazard!

⇒ The conditional failure rate at any time y, given that the event has not occurred prior to time y, does not depend upon y.

⇒ Called "no-aging" property or the "old as good as new" property.

⋆⋆ In other words, lack of memory property!

$$\Pr(Y \geq y + z \mid Y \geq y) = \Pr(Y \geq z)$$

⇒ The time until the future occurrence of an event does not depend upon past history.

⇒ mathematical tractability but too restrictive in applications, especially for health and industrial applications.

y: survival time

- Likelihood under the exponential distribution

Given $\mathcal{D} = (n, \widetilde{\mathbf{y}}, \boldsymbol{\nu})$ $\tilde{y} = \min(y, c)$

$$\mathcal{L}(\lambda \mid \mathcal{D}) = \prod_{i=1}^{n} f(\widetilde{y}_i \mid \lambda)^{\nu_i} S(\widetilde{y}_i \mid \lambda)^{(1-\nu_i)}$$

$$= \prod_{i=1}^{n} (\lambda e^{-\lambda \tilde{y}_i})^{\nu_i} (e^{-\lambda \tilde{y}_i})^{1-\nu_i}$$

$$= \lambda^{\Sigma \nu_i} e^{-\lambda \Sigma \tilde{y}_i}$$

- Prior for λ

  ⋆⋆ The conjugate prior for λ is the gamma prior, $\lambda \sim \text{Gamma}(\alpha_0, \lambda_0)$;

$$\pi(\lambda \mid \alpha_0, \lambda_0) \propto \lambda^{\alpha_0 - 1} \exp(-\lambda_0 \lambda).$$

  ⋆⋆ The posterior of λ is given by

$$\pi(\lambda \mid D) \propto \underbrace{\lambda^{\alpha_0 - 1} e^{-\lambda_0 \lambda}}_{\text{prior}} \underbrace{\lambda^{\Sigma \nu_i} e^{-\lambda \Sigma \hat{y}_i}}_{\text{likelihood}}$$

$$= \lambda^{(\alpha_0 + \Sigma \nu_i) - 1} e^{-\lambda(\lambda_0 + \Sigma \hat{y}_i)}$$

$\Rightarrow$ $\pi(\lambda \mid D)$ is $\text{Ga}(\alpha_0 + \Sigma \nu_i, \lambda_0 + \Sigma \hat{y}_i)$

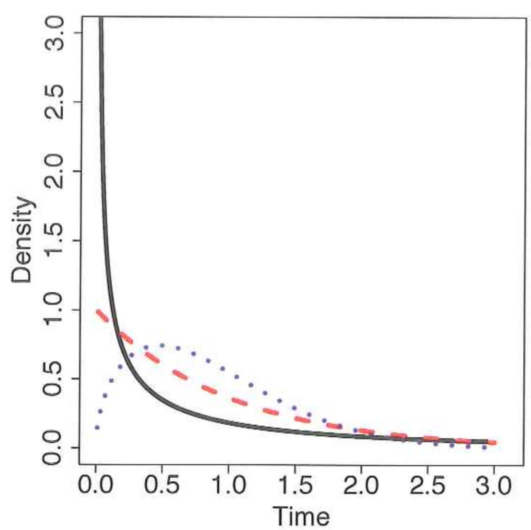
* Check ICS 2.1 for the predictive distribution.

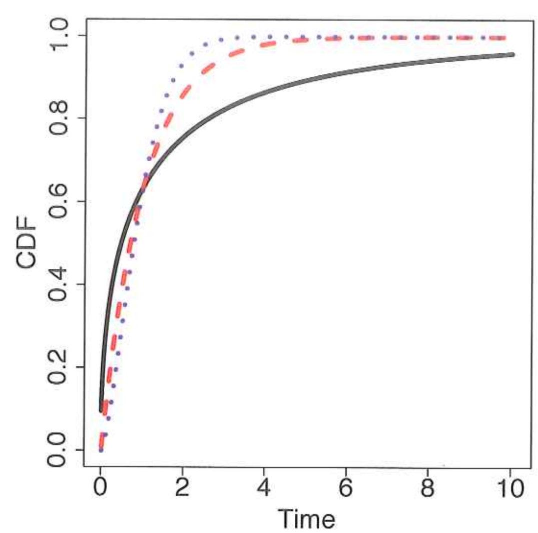
• Fixed λ
• Vary α $\Rightarrow$ show how α affects pdf & cdf

♢ eg.: $T \sim$ Weibull(α, λ) ($\lambda > 0$ and $\alpha > 0$). That is,

$$f(t) = \begin{cases} \alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha}), & \text{for } t > 0, \\ 0 & \text{otherwise .} \end{cases}$$

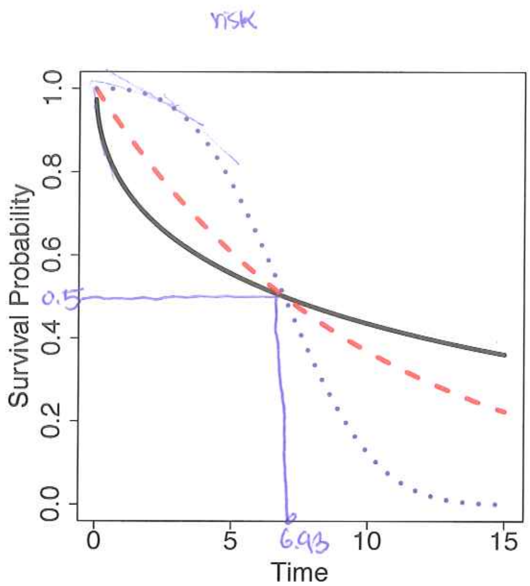
⋆⋆ $(\lambda, \alpha) = (1, 0.5)$ for black solid, (1, 1) for red dashed and (1, 1.5) for blue dotted.

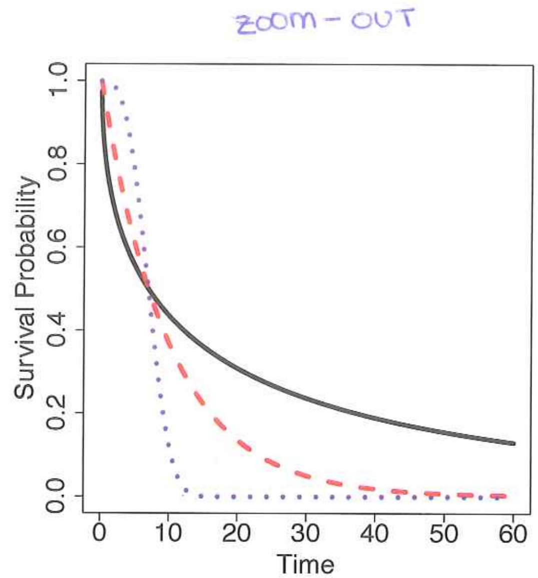




$6.93 = \overset{\parallel}{F(t)}$

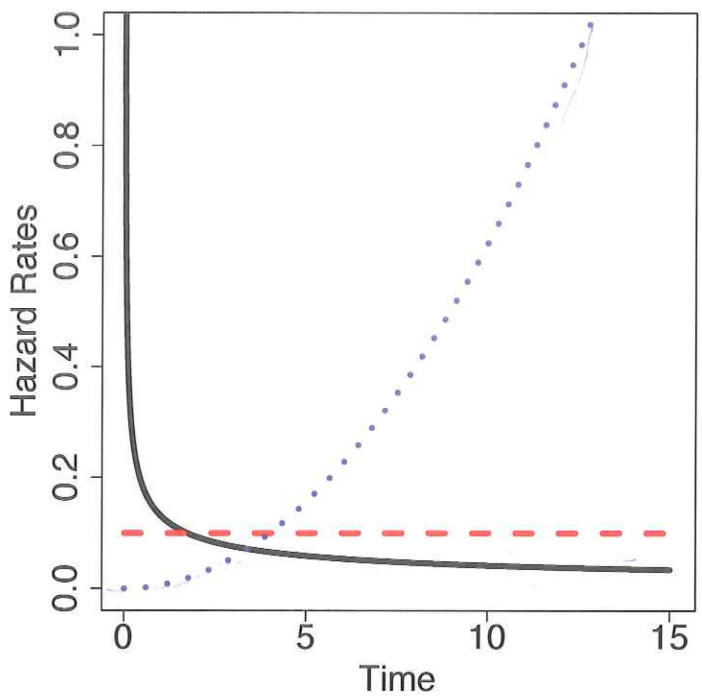
◇ eg.: $T \sim \text{Weibull}(\alpha, \lambda)$ ($\alpha > 0$ and $\lambda > 0$) $\Rightarrow$ $S(t) = \exp(-\lambda t^{\alpha})$. Survival curves with a common median of 6.93.





⋆⋆ $(\alpha, \lambda) = (0.5, 0.26328)$ for black solid, (1, 0.1) for red dashed and (3, 0.00208) for blue dotted.

$\diamondsuit$ eg.: $T \sim \mathrm{Weibull}(\alpha, \lambda)$ ($\alpha > 0$ and $\lambda > 0$) $\Rightarrow h(t) = \alpha\lambda t^{\alpha-1}$.



$\star\star$: $(\alpha, \lambda) = (0.5, 0.26328)$ for black solid, (1, 0.1) for red dashed and (3, 0.00208) for blue dotted.

♣ **Weibull** distribution:

$$y_i \mid \alpha, \gamma \overset{iid}{\sim} \mathrm{W}(\alpha, \gamma),$$

where scale parameter $\gamma > 0$ and shape parameter $\alpha > 0$.

$= \alpha \exp((\alpha-1)\log y + \lambda - \exp(\lambda) y^{\alpha})$

- Density function $= \alpha y^{\alpha-1} \exp(\lambda - \exp(\lambda) y^{\alpha})$ $\quad \lambda = \log \gamma$

$\Rightarrow \gamma = e^{\lambda}$

$$f(y \mid \alpha, \gamma) = \begin{cases} \gamma \alpha y^{\alpha-1} \exp(-\gamma y^{\alpha}) & \text{if } y > 0, \\ 0, & \text{o.w.} \end{cases}$$

- **Caution!** KM uses ~~a~~ different letters and Wiki uses a different parameterization (as I said before).
- ICS uses a different parameterization by letting $\lambda = \log(\gamma)$ (so, $-\infty < \lambda < \infty$).
- Survival function: $S(y \mid \alpha, \gamma) = \exp(-\gamma y^{\alpha})$.
- The exponential distribution is a special case when $\alpha = 1$.

- The *most popular* parametric survival model.
- Why so popular? Well.... among many reasons...

⋆⋆ The hazard function $h(y \mid \alpha, \gamma) = \gamma \alpha y^{\alpha-1}$

- $\alpha > 1$: increasing hazard
- $\alpha = 1$: constant hazard
- $\alpha < 1$: decreasing hazard

- Important generalization of the exponential distribution; allows for dependence of the hazard on time.
- Also, it has relatively simple survival and pdf.
- We already looked at many properties of the Weibull distribution in Lec 1.

- Likelihood under the Weibull distribution

Given $\mathcal{D} = (n, \widetilde{\mathbf{y}}, \boldsymbol{\nu})$ $\quad \tilde{y} = \min(y, c)$

$$\mathcal{L}(\alpha, \lambda \mid \mathcal{D}) = \prod_{i=1}^{n} f(\widetilde{y}_i \mid \alpha, \lambda)^{\nu_i} S(\widetilde{y}_i \mid \alpha, \lambda)^{(1-\nu_i)}$$

$$= \prod_{i=1}^{n} \left\{ \alpha \exp\left((\alpha-1)\log(\tilde{y}_i) + \lambda - \exp(\lambda)\tilde{y}_i^{\alpha} \right) \right\}^{\nu_i}$$

$$\times \left\{ \exp\left(-\exp(\lambda)\tilde{y}_i^{\alpha} \right) \right\}^{1-\nu_i}$$

$$= \alpha^{\Sigma \nu_i} \exp\left[\sum_i \left\{ \nu_i (\alpha-1)\log(\tilde{y}_i) + \nu_i \lambda - \exp(\lambda) y_i^{\alpha} \right\} \right]$$

If α is known, the likelihood for $\gamma = e^{\lambda}$ is

$$f(\gamma \mid D) = \gamma^{\Sigma \nu_i} \exp\left(-\gamma \Sigma y_i^{\alpha} \right)$$

$\Rightarrow$ the conjugate prior for γ is Gea

- Recall

$$f(y \mid \alpha, \gamma) = \gamma\alpha y^{\alpha-1} \exp(-\gamma y^{\alpha})$$

⋆⋆ Prior for γ with known α: let $\gamma \sim \text{Gamma}(a, b)$ (conjugate!)

⋆⋆ No joint conjugate prior for α and γ

⋆⋆ Joint prior for α and γ: assume a priori independence

$$\alpha \sim \text{Gamma}(\alpha_0, \kappa_0), \qquad \text{and} \qquad \lambda = \log(\gamma) \sim N(\mu_0, \sigma_0^2).$$

$\underline{\alpha > 0}$ $\qquad\qquad\qquad\qquad$ $\lambda \in \mathbb{R}$

$$\pi(\alpha, \lambda \mid D) \propto \mathcal{L}(\alpha, \lambda \mid D)\, \pi(\alpha)\, \pi(\lambda)$$

$$\propto \underline{\alpha^{\Sigma \nu_i}} \exp\left\{ \sum_{i=1}^{n} \left(\nu_i \lambda + \nu_i(\alpha-1) \cdot \log(\tilde{y}_i) - \exp(\lambda)\tilde{y}_i^{\alpha} \right) \right\}$$

$$\times \underline{\alpha^{\alpha_0 - 1}} \exp(\kappa_0 \alpha) \cdot \exp\left(-\frac{1}{2\sigma_0^2}(\lambda - \mu_0)^2\right)$$

No closed form $\Rightarrow$ Gibbs Sampling

① $\pi(\alpha \mid \lambda, D) \propto \alpha^{\Sigma \nu_i + \alpha_0 - 1} \exp\left\{ \sum_{i=1}^{n} \left(\nu_i(\alpha-1)\log(\tilde{y}_i) - \exp(\lambda)\tilde{y}_i^{\alpha} \right) + \kappa_0 \alpha \right\}$

② $\pi(\lambda \mid \alpha, D) \propto \exp\left\{ \sum_{i=1}^{n} \left(\nu_i \lambda + \nu_i(\alpha-1)\log(\tilde{y}_i) - \exp(\lambda)\tilde{y}_i^{\alpha} \right) - \frac{(\lambda - \mu_0)^2}{2\sigma_0^2} \right\}$

- Recall

$$\mathcal{L}(\alpha, \lambda \mid \mathcal{D}) = \alpha^{\sum \nu_i} \exp\left\{\sum (\nu_i \lambda + \nu(\alpha - 1)\log(y_i) - \exp(\lambda) y_i^{\alpha})\right\}.$$

⋆⋆ Joint prior for α and γ: assume a priori independence

$$\alpha \sim \text{Gamma}(\alpha_0, \kappa_0), \qquad \text{and} \qquad \lambda = \log(\gamma) \sim \text{N}(\mu_0, \sigma_0^2).$$

- The posterior of α and λ is given by

◦ Type I extreme value distribution
◦ Gumbel distribution

♣ **Extreme Value** distribution: $-\infty < y_i < \infty$

$$y_i \mid \alpha, \lambda \overset{iid}{\sim} \mathrm{V}(\alpha, \lambda), \quad \alpha > 0 \text{ and } -\infty < \lambda = \log(\gamma) < \infty.$$

- Connection to the Weibull distribution:
  - ⋆⋆ Sometimes it is useful to work with the logarithm of the lifetimes. Suppose $T \mid \alpha, \lambda \sim W(\alpha, \lambda)$.
  - ⋆⋆ We can check $Y = \log(T)$ follows the extreme value distribution.

- The density function is

$$f(y \mid \alpha, \lambda) = \alpha \exp(\underbrace{\lambda + \alpha y} - \exp(\lambda + \alpha y)), \quad -\infty < y < \infty.$$

- Writing the model in a general linear model format, $\underline{y = \mu + \sigma w}$.

let $\underbrace{\mu = -\frac{\lambda}{\alpha}}_{\text{location parameter}}$ & $\underbrace{\sigma = \frac{1}{\alpha}}_{\text{scale parameter}}$ $\Rightarrow y = -\frac{\lambda}{\alpha} + \frac{1}{\alpha} W$

$\Leftrightarrow W = \lambda + \alpha y$

$$f(w) = \cancel{\alpha} \exp(W - \exp(w)) \times \frac{1}{\cancel{\alpha}}$$

$$= \exp(W - \exp(w)), \quad -\infty < W < \infty$$

$\Rightarrow$ called the standard extreme value distribution.

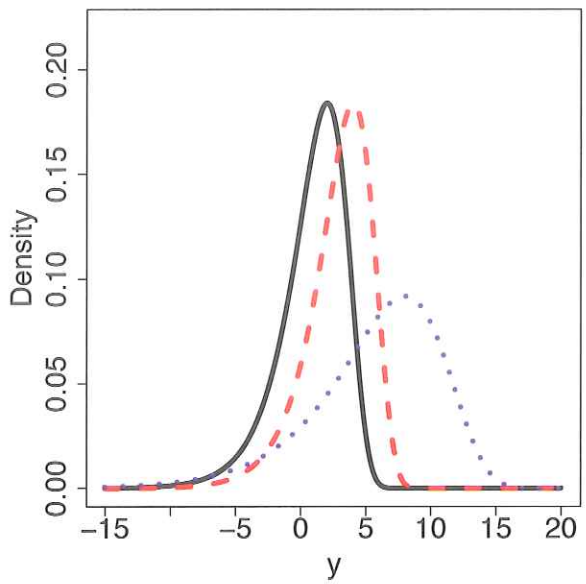
* Why do we discuss this?

- Replace μ with $X\beta$ where X: covariate

$\Rightarrow$ relate ~~covaria~~ survival time to covariates.

$\boxed{\text{Lec 3}}$

- Density of V(α, λ)



$\star\star$: $(\alpha, \lambda) = (-1, 0.5)$ for black solid, (-2.0, 0.5) for red dashed and (-2.0, 0.55) for blue dotted.

- The survival function is

$$S(y \mid \alpha, \lambda) = \exp(-\exp(\lambda + \alpha y)).$$

- Likelihood under extreme value distribution

Given $\mathcal{D} = (n, \tilde{\mathbf{y}}, \boldsymbol{\nu})$

$$\mathcal{L}(\alpha, \lambda \mid \mathcal{D}) = \prod_{i=1}^{n} f(\tilde{y}_i \mid \alpha, \lambda)^{\nu_i} S(\tilde{y}_i \mid \alpha, \lambda)^{(1-\nu_i)}$$

$$= \alpha^{\Sigma \nu_i} \exp\left[\sum_{i=1}^{n} \left\{ \nu_i(\lambda + \alpha y_i) - \exp(\lambda + \alpha y_i) \right\} \right]$$

- **Extreme Value**: The density function

$$f(y \mid \alpha, \lambda) = \alpha \exp(\lambda + \alpha y - \exp(\lambda + \alpha y)), \quad -\infty < y < \infty.$$

⋆⋆ No joint conjugate prior for α and λ

⋆⋆ Joint prior for α and λ: assume a priori independence

$$\alpha \sim \text{Gamma}(\alpha_0, \kappa_0), \qquad \text{and} \qquad \lambda = \log(\gamma) \sim \text{N}(\mu_0, \sigma_0^2).$$

- The posterior of α and λ is given by

$$\pi(\alpha, \lambda \mid D) \propto \alpha^{\sum \nu_i} \exp\left(\sum \{ \nu_i(\lambda + \alpha y_i) - \exp(\lambda + \alpha y_i) \} \right)$$
$$\times\ \alpha^{\alpha_0 - 1} \exp(-\kappa_0 \alpha) \cdot \exp\left(-\frac{(\lambda - \mu_0)^2}{2\sigma_0^2}\right)$$

$\Rightarrow$ need Gibbs Sampling

$\Rightarrow$ need to find the full conditionals

♣ So far we have discussed:

$y = \log(t)$

Original Scale $t > 0$	Log Scale $-\infty < y < \infty$
Weibull	Extreme Value
Log Normal	Normal

♣ **Log Normal** distribution:

$$y_i \mid \mu, \sigma^2 \overset{iid}{\sim} \text{LN}(\mu, \sigma^2), \quad -\infty < \mu < \infty \text{ and } \sigma^2 > 0.$$

- Connection to the normal distribution: $\log(y) \sim \text{N}(\mu, \sigma^2)$.
- Another very popular model!

$\Rightarrow \log(y) = \mu + \sigma W$

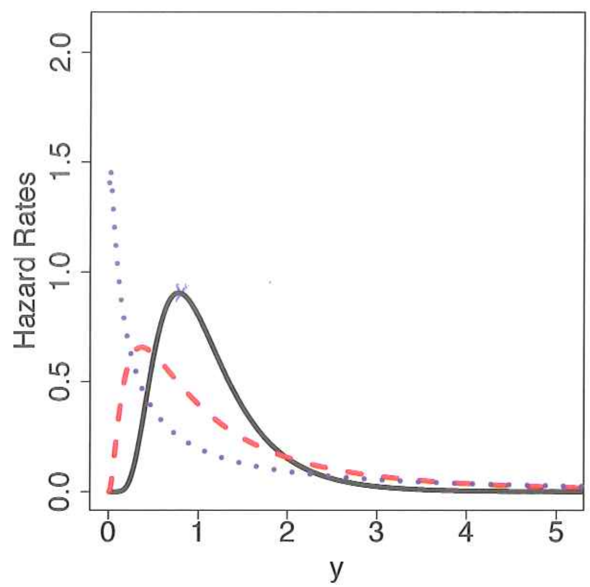
$W \sim N(0, 1)$

- The density function is

$$f(y \mid \mu, \sigma) = \frac{\exp\{-\frac{1}{2}(\frac{\log(y)-\mu}{\sigma})^2\}}{y(2\pi)^{1/2}\sigma} = \frac{1}{y} \cdot \phi\left(\frac{\log(y) - \mu}{\sigma}\right),$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution.

- Densities of Log Normal.



⋆⋆: $(\mu, \sigma) = (0, 0.5)$ for black solid, (0, 1) for red dashed and (0, 2) for blue dotted.

$= P(Y > y \mid \mu, \sigma^2)$

$= P(\underbrace{\log(Y)} > \log(y) \mid \mu, \sigma^2)$

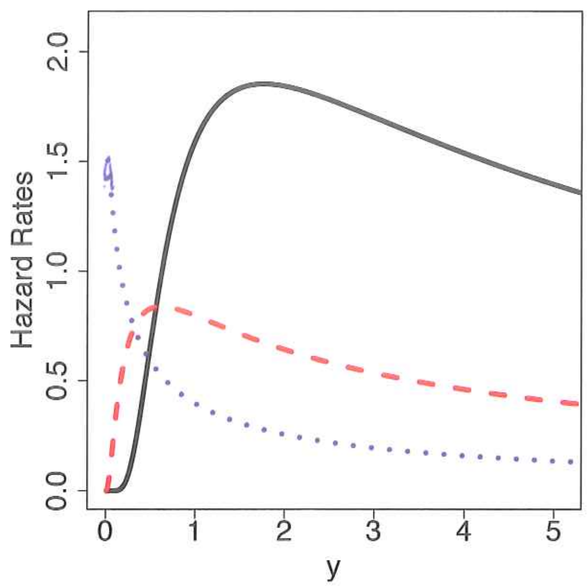
- The survival function is

$$S(y \mid \mu, \sigma^2) = 1 - \Phi(\frac{\log(y) - \mu}{\sigma}),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

- The hazard rate is $f(y \mid \mu, \sigma^2)/S(y \mid \mu, \sigma^2)$: Hump-shaped!
  - Its value at 0 is zero.
  - It increase to a maximum and, then, decreases to 0 as y approaches ∞.
  - May be not appropriate as a lifetime distribution but still good when large values of y are not of interest.

- The hazard rate is $f(y \mid \mu, \sigma^2)/S(y \mid \mu, \sigma^2)$.



★★: $(\mu, \sigma) = (0, 0.5)$ for black solid, (0, 1) for red dashed and (0, 2) for blue dotted.

- **Log Normal** distribution: We have

$$f(y \mid \mu, \sigma^2) = \frac{\exp\{-\frac{1}{2}(\frac{\log(y)-\mu}{\sigma})^2\}}{y(2\pi)^{1/2}\sigma} \quad \& \quad S(y \mid \mu, \sigma^2) = 1 - \Phi(\frac{\log(y)-\mu}{\sigma}).$$

  ⋆⋆ No joint conjugate prior for μ and σ^2

  ⋆⋆ Let $\tau = 1/\sigma^2$. Joint prior for μ and τ: ~~assume a priori independence~~

$$\mu \sim \mathrm{N}(\mu_0, \frac{1}{\tau_0\ \tau}), \qquad \text{and} \qquad \tau \sim \mathrm{Gamma}(\alpha_0/2, \lambda_0/2).$$

- The posterior of μ and τ is given by

$$\pi(\mu, \tau \mid D) \propto \prod_{i=1}^{n} \left\{ \frac{\tau^{\frac{1}{2}}}{\sqrt{2\pi}\,\tilde{y}_i} \exp\left(-\frac{\tau}{2}(\log(\tilde{y}_i) - \mu)^2\right) \right\}^{\nu_i}$$

$$\times \left\{ 1 - \Phi\left(\tau^{1/2}(\log(\tilde{y}_i) - \mu)\right) \right\}^{1-\nu_i}$$

$$(\tau_0\tau)^{\frac{1}{2}} \exp\left(-\frac{\tau\tau_0(\mu-\mu_0)^2}{2}\right) \tau^{\frac{\alpha_0}{2}-1}\, e^{-\frac{\tau\lambda_0}{2}}$$

$$\propto \tau^{\frac{\alpha_0 + \Sigma \nu_i - 1}{2}} \exp\left\{-\frac{\tau}{2}\left(\Sigma \nu_i (\log(\tilde{y}_i) - \mu)^2 + \tau_0(\mu-\mu_0)^2 + \lambda_0\right)\right\}$$

$$\times \prod_{i=1}^{n} \frac{1}{\tilde{y}_i^{\nu_i}} \left(1 - \Phi(\tau(\log(y_i) - \mu))\right)^{1-\nu_i}$$