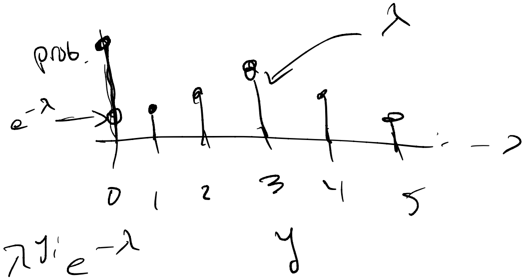
Excess zeros:

We need a way to account for more zeros in data than can be accommodated by a standard distn.

For example: Suppose y_i, $i=1,\dots,n$ are non-negative integers with empirical pmf:



If $y_i \sim \text{Pois}(\lambda)$

Then $p(y_i=0 \mid \lambda) = \dfrac{\lambda^{y_i} e^{-\lambda}}{y_i!}$

$$= e^{-\lambda}$$

Thus, if $\lambda = 3$ then $e^{-3} \approx 0.05$ and $P(y_i = 3 \mid \lambda) = 0.22$

The Poisson can't handle that many zeros!

Modeling options:

1.) Hurdle model:

$$y_i \sim \begin{cases} \text{zero} \\ \text{distn not incl. zero} \end{cases}$$

can inflate and deflate zeros

2.) zero-inflated model:

$$y_i \sim \begin{cases} \text{zero} \\ \text{distn incl. zero} \end{cases}$$

can only inflate zeros.

Scenario: chip manufacturing plant checks for # of defects y_i on each chip.

a.) Two checkers: 1.) yes/no 2.) if yes, count defects. Both checkers are perfect.

$\Rightarrow$ Model: hurdle

b.) First checker is imperfect.

$\Rightarrow$ Model: zero-inflated

Zero-inflated Poisson:

$$y_i \sim \begin{cases} 1_{\{y_i=0\}} & , \ z_i = 0 \\ \text{Pois}(\lambda) & , \ z_i = 1 \end{cases} \quad , \ i = 1, \dots, n$$

$$z_i \sim \text{Bern}(p)$$
$$\lambda \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda)$$
$$p \sim \text{Beta}(\alpha_p, \beta_p)$$

Posterior:

$$[\underline{z}, \lambda, p \mid y] \propto \left(\prod_{i=1}^{n} [y_i \mid \lambda]^{z_i} \, 1_{\{y_i=0\}}^{1-z_i} [z_i \mid p] \right) [\lambda][p]$$

Full-conditionals:

(from before)

$$[z_i \mid \cdot] = \text{Bern}(\tilde{p}_i), \quad \tilde{p}_i = \frac{p [y_i \mid \lambda]}{p[y_i \mid \lambda] + (1-p) 1_{\{y_i=0\}}}$$

$\Rightarrow$ when $y_i > 0 \Rightarrow z_i = 1$

and when $y_i = 0 \Rightarrow \tilde{p}_i = \dfrac{p e^{-\lambda}}{p e^{-\lambda} + 1 - p}$

$$[p \mid \cdot] = \text{Beta}\left(\sum_{i=1}^{n} z_i + \alpha_p \, , \, \sum_{i=1}^{n} (1-z_i) + \beta_p\right) \quad \left(\text{from before}\right)$$

$$[\lambda \mid \cdot] \propto \left(\prod_{i=1}^{n} [y_i \mid \lambda]^{z_i}\right)[\lambda]$$

$$\propto \lambda^{\sum_{\forall z_i = 1} y_i} \, e^{-\left(\sum_{i=1}^{n} z_i\right)\lambda} \, \lambda^{\alpha_\lambda + 1} \, e^{-\beta_\lambda \lambda}$$

$$\propto \lambda^{\underbrace{\sum_{\forall z_i = 1} y_i + \alpha_\lambda + 1}_{\tilde{\alpha}}} \, e^{-\underbrace{\left(\sum_{i=1}^{n} z_i + \beta_\lambda\right)}_{\tilde{\beta}}\lambda}$$

$$= \text{Gamma}(\tilde{\alpha} \, , \, \tilde{\beta})$$

If the count data are zero-inflated and overdispersed relative to the Poisson, we can use a zero-inflated negative binomial model:

$$y_i \sim \begin{cases} NB(\lambda, N), & z_i = 1 \\ 1_{\{y_i = 0\}}, & z_i = 0 \end{cases}$$

$$z_i \sim \text{Bern}(p)$$
$$p \sim \text{Beta}(\alpha_p, \beta_p)$$
$$\log \lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$$
$$\log N \sim N(\mu_N, \sigma_N^2)$$

Posterior:

$$[\mathbf{z}, p, \log\lambda, \log N \mid \mathbf{y}] \propto \prod_{i=1}^{n} \left([y_i \mid \lambda, N]^{z_i} 1_{\{y_i=0\}}^{1-z_i} \times [z_i \mid p] \right) [p][\log\lambda][\log N]$$

Full-conditional Distributions:

$[p \mid \cdot] \propto$ same as for ZIP

$[z_i \mid \cdot] = \text{Bern}(\tilde{p}_i)$, $\tilde{p}_i = \dfrac{p[y_i \mid \lambda, N]}{p[y_i \mid \lambda, N] + 1 - p}$

when $y_i = 0$, else let $z_i = 1$.

$$= \frac{p\left(\frac{N}{N+\lambda}\right)^N}{p\left(\frac{N}{N+\lambda}\right)^N + 1 - p}$$

$$\left.[\log \lambda \mid \cdot] \propto \prod_{i=1}^{n} [y_i \mid \lambda, N]^{z_i} [\log \lambda]\right\}$$ use M-H

$$\left.[\log N \mid \cdot] \propto \prod_{i=1}^{n} [y_i \mid \lambda, N]^{z_i} [\log N]\right\}$$ use M-H