Six types of size-bias

Matthew Aldridge

School of Mathematics, University of Leeds

Thursday 16 October 2024

Section 1

Size-bias

Houses in Leeds

I want to find out how many people live in different houses in Leeds.

Houses in Leeds

I want to find out how many people live in different houses in Leeds.

I want to estimate:

- $p(0)$, the proportion of houses that are uninhabited
- $p(1)$, the proportion of houses with 1 person living in them
- $p(2)$, the proportion of houses with 2 people living in them
- $p(3)$, the proportion of houses with 3 people living in them
- etc.

Asking people about houses

I want to find out how many people live in different houses in Leeds.

I want to estimate:

- $p(x)$, the proportion of houses with x people living in them

I will go out into the city centre and ask people:

"How many people live in your house?"

Asking people about houses

I want to find out how many people live in different houses in Leeds.

I want to estimate:

- $p(x)$, the proportion of houses with x people living in them

I will go out into the city centre and ask people:

"How many people live in your house?"

Question:

What mistake have I made?

My mistake

I will go out into the city centre and ask people:

"How many people live in your house?"

Question:

What mistake have I made?

- I can never find out about the uninhabited houses:
  there's no one living there to ask!

My mistake

I will go out into the city centre and ask people:

"How many people live in your house?"

Question:

What mistake have I made?

- I can never find out about the uninhabited houses: there's no one living there to ask!
- Houses with lots of people living in them will be overrepresented:
  - A townhouse in Hyde Park with 8 students packed in gives me 8 opportunities to bump into one of them.
  - A flat by the station with 1 young professional living alone gives me only 1 chance to survey him.

What am I *actually* counting?

I'm actually finding not

$$p(x) = \text{proportion of houses with } x \text{ people}$$

but rather

$$p^*(x) = \text{proportion of people living in a house with } x \text{ people.}$$

What am I *actually* counting?

I'm actually finding not

$$p(x) = \text{proportion of houses with } x \text{ people}$$

but rather

$$p^*(x) = \text{proportion of people living in a house with } x \text{ people.}$$

$$p^*(x) \propto x\, p(x)$$

What am I *actually* counting?

I'm actually finding not

$$p(x) = \text{proportion of houses with } x \text{ people}$$

but rather

$$p^*(x) = \text{proportion of people living in a house with } x \text{ people.}$$

$$p^*(x) \propto x\, p(x)$$

$$p^*(x) = \frac{1}{\sum_y y\, p(y)}\, x\, p(x) = \frac{1}{\mu}\, x\, p(x),$$

where $\mu = \sum_y y\, p(y) = \mathbb{E}X$.

Class sizes

I am the module manager for a class of 30 students.

For tutorials, they are split into three unequal groups of size 6, 11, and 13.

Question:

What is the distribution of tutorial group sizes?

Class sizes

I am the module manager for a class of 30 students.

For tutorials, they are split into three unequal groups of size 6, 11, and 13.

Question:

What is the distribution of tutorial group sizes?

From **my point of view**:

$$p(6) = \tfrac{1}{3} \qquad p(11) = \tfrac{1}{3} \qquad p(13) = \tfrac{1}{3}$$

Class sizes

I am the module manager for a class of 30 students.

For tutorials, they are split into three unequal groups of size 6, 11, and 13.

Question:

What is the distribution of tutorial group sizes?

From **my point of view**:

$$p(6) = \tfrac{1}{3} \qquad p(11) = \tfrac{1}{3} \qquad p(13) = \tfrac{1}{3}$$

From **the students' point of view**:

$$p^*(6) = \tfrac{6}{30} \qquad p^*(11) = \tfrac{11}{30} \qquad p^*(13) = \tfrac{13}{30}$$

Class sizes

I am the module manager for a class of 30 students.

For tutorials, they are split into three unequal groups of size 6, 11, and 13.

Question:

What is the distribution of tutorial group sizes?

From **my point of view**:

$$p(6) = \tfrac{1}{3} \qquad p(11) = \tfrac{1}{3} \qquad p(13) = \tfrac{1}{3}$$

From **the students' point of view**:

$$p^*(6) = \tfrac{6}{30} \qquad p^*(11) = \tfrac{11}{30} \qquad p^*(13) = \tfrac{13}{30}$$

Again, $p^*(x) = \dfrac{1}{\mu}\, x\, p(x)$.

Size-bias: first definition

Definition (first attempt)

Let X be a random variable on the non-negative integers $\mathbb{N}$
with PMF p and expectation $\mu = \mathbb{E}X$.
Then the random variable X^*
with PMF p^*, where

$$p^*(x) = \frac{1}{\mu}\, x\, p(x),$$

is called *the size-biased version of* X, or just the *size-bias* of X.

Size-bias: first definition

Definition (first attempt)

Let X be a random variable on the non-negative integers $\mathbb{N}$ with PMF p and expectation $\mu = \mathbb{E}X$.
Then the random variable X^* with PMF p^*, where

$$p^*(x) = \frac{1}{\mu}\, x\, p(x),$$

is called *the size-biased version of* X, or just the *size-bias* of X.

But we'd like a definition that works equally well for continuous random variables with PDF f

$$f^*(x) = \frac{1}{\mu}\, x\, f(x),$$

and for random variables that are neither discrete nor purely continuous, etc. . .

Size-bias: better definition

A better definition is this:

Definition

Let X be a random variable with expectation $\mu = \mathbb{E}X$.
The *size-biased version of* X, or just the *size-bias* of X, is the random variable X^* such that

$$\mathbb{E}\, X\, f(X) = \mu\, \mathbb{E}\, f(X^*)$$

for all "nice" functions f.

Taking f to be an appropriate indicator function recovers the previous definitions.

Size-bias: existence

Definition

Let X be a random variable with expectation $\mu = \mathbb{E}X$.
The *size-biased version of* X, or just the *size-bias* of X, is the random variable X^* such that

$$\mathbb{E}\, X\, f(X) = \mu\, \mathbb{E}\, f(X^*)$$

for all "nice" functions f.

The size-bias X^* exists and is unique provided:

- X is almost surely non-negative (or non-positive)
- X has finite expectation
- X is not a point mass at 0

Examples of size-bias

Examples

Distribution	**Size-bias on** j
$\Gamma(\nu, \lambda)$	$\Gamma(\nu+1, \lambda)$
$\mathrm{Beta}(\alpha, \beta)$	$\mathrm{Beta}(\alpha+1, \beta)$
point mass at x	point mass at x

The point mass is the only case where X and X^* have the same distribution.

Properties of size-bias

Theorem

Let X be a non-negative random variable with finite expectation $\mathbb{E}X = \mu$ and let X^* be its size-bias.

1. *Expectation:* $\mathbb{E}X^* = \mu + \dfrac{\mathrm{Var}(X)}{\mu}$
2. *Moments:* $\mathbb{E}(X^*)^k = \dfrac{1}{\mu}\,\mathbb{E}X^{k+1}$
3. *MGF:* $M_{X^*}(t) = \dfrac{1}{\mu}\dfrac{\mathrm{d}}{\mathrm{d}t}M_X(t)$
4. *Size-bias commutes with scaling:* $(aX)^* \equiv a(X^*)$

Properties of size-bias: proof

Theorem

1. *Expectation:* $\mathbb{E}X^* = \mu + \dfrac{\mathrm{Var}(X)}{\mu}$
2. *Moments:* $\mathbb{E}(X^*)^k = \dfrac{1}{\mu}\,\mathbb{E}X^{k+1}$
3. *MGF:* $M_{X^*}(t) = \dfrac{1}{\mu}\,\dfrac{\mathrm{d}}{\mathrm{d}t}\,M_X(t)$
4. *Size-bias commutes with scaling:* $(aX)^* \equiv a(X^*)$

Proof.

In the definition $\mathbb{E}\,X\,f(X) = \mu\,\mathbb{E}\,f(X^*)$,

1. take $f(x) = x$
2. take $f(x) = x^k$
3. take $f(x) = \mathrm{e}^{tx}$

Point 4 follows from point 3, since $M_{aX}(t) = M_X(at)$. $\square$

Section 2

Reduced size-bias

My alternative definition

Definition

Let X be a random variable with expectation $\mu = \mathbb{E}X$.
The *size-biased version of* X, or just the *size-bias* of X,
is the random variable X^* such that

$$\mathbb{E}\, X\, f(X) = \mu\, \mathbb{E}\, f(X^*).$$

For discrete random variables on the non-negative integers, I propose . . .

My alternative definition

Definition

Let X be a random variable with expectation $\mu = \mathbb{E}X$.
The *size-biased version of* X, or just the *size-bias* of X, is the random variable X^* such that

$$\mathbb{E}\, X\, f(X) = \mu\, \mathbb{E}\, f(X^*).$$

For discrete random variables on the non-negative integers, I propose . . .

Definition

Let X be a discrete random variable with expectation $\mu = \mathbb{E}X$.
The *reduced size-biased version of* X, or just the *reduced size-bias* of X, is the discrete random variable $X^!$ such that

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^!).$$

. . . is a better definition.

Two views of reduced size bias

Definition

Let X be a discrete random variable with expectation $\mu = \mathbb{E}X$.
The *reduced size-biased version of* X, or just the *reduced size-bias* of X, is the discrete random variable $X^!$ such that

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^!)$$

for all "nice" functions f.

I want to explain why I think reduced size-bias...

1. ...is just a minor notational change of convention compared to the usual size-bias,

Two views of reduced size bias

Definition

Let X be a discrete random variable with expectation $\mu = \mathbb{E}X$.
The *reduced size-biased version of* X, or just the *reduced size-bias* of X, is the discrete random variable $X^!$ such that

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^!)$$

for all "nice" functions f.

I want to explain why I think reduced size-bias...

1. ... is just a minor notational change of convention compared to the usual size-bias,
2. ... has surprising and beneficial properties that are not obvious from the minor change of convention.

Reduced size bias: PMF

Definition

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^{!})$$

Setting f to be the indicator function at x, we get

$$p^{!}(x) = \frac{1}{\mu}(x+1)\, p(x+1)$$

Reduced size bias: PMF

Definition

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^!)$$

Setting f to be the indicator function at x, we get

$$p^!(x) = \frac{1}{\mu}(x+1)\, p(x+1)$$

...compared to

$$p^*(x) = \frac{1}{\mu}\, x\, p(x).$$

Reduced size bias: PMF

Definition

$$\mathbb{E}\, X\, f(X-1) = \mu\, \mathbb{E}\, f(X^!)$$

Setting f to be the indicator function at x, we get

$$p^!(x) = \frac{1}{\mu}(x+1)\, p(x+1)$$

…compared to

$$p^*(x) = \frac{1}{\mu}\, x\, p(x).$$

So $p^!(x) = p^*(x+1)$, and $X^! = X^* - 1$.

The reduced size-bias is just the usual size-bias minus 1.

Examples of reduced size-bias

Examples

Distribution	**Size-bias**	**Reduced size-bias**
$\mathrm{Bin}(n, p)$	$\mathrm{Bin}(n-1, p) + 1$	$\mathrm{Bin}(n-1, p)$
$\mathrm{NegBin}(n, p)$	$\mathrm{NegBin}(n+1, p) + 1$	$\mathrm{NegBin}(n+1, p)$
$\mathrm{Po}(\lambda)$	$\mathrm{Po}(\lambda) + 1$	$\mathrm{Po}(\lambda)$
$\mathrm{Herm}(\mu, \sigma^2)$	$\mathrm{Herm}(\mu, \sigma^2) + \mathrm{Bern}(\frac{\sigma^2}{\mu}) + 1$	$\mathrm{Herm}(\mu, \sigma^2) + \mathrm{Bern}(\frac{\sigma^2}{\mu})$

The Hermite distribution $X \sim \mathrm{Herm}(\mu, \sigma^2)$ is
$X = Y + 2Z$ where $Y \sim \mathrm{Po}(\mu - \sigma^2)$ and $Z \sim \mathrm{Po}(\sigma^2/2)$.

Properties of reduced size-bias

Theorem

Let X be a discrete random variable with finite expectation $\mathbb{E}X = \mu$ and let $X^!$ be its reduced size-bias.

1. *Expectation:* $\mathbb{E}X^! = \mu + \dfrac{\mathrm{Disp}(X)}{\mu}$
2. *Factorial moments:* $\mathbb{E}(X^!)^{\underline{k}} = \dfrac{1}{\mu}\,\mathbb{E}X^{\underline{k+1}}$
3. *PGF:* $G_{X^!}(t) = \dfrac{1}{\mu}\,\dfrac{\mathrm{d}}{\mathrm{d}t}\,G_X(t)$
4. *Reduced size-bias commutes with thinning:* $(a \circ X)^! \equiv a \circ (X^!)$

- Dispersion: $\mathrm{Disp}(X) = \mathrm{Var}(X) - \mathbb{E}X$
- Factorial moments: $\mathbb{E}\,X^{\underline{k}} = \mathbb{E}\,X(X-1)\cdots(X-k+1)$
- Thinning: $a \circ X \sim \mathrm{Bin}(X, a)$, for $a \in [0,1]$.

Poisson as fixed point

Theorem

A discrete random variable X and its reduced size-bias $X^!$ have the same distribution if and only if X is a Poisson distribution

Poisson as fixed point

Theorem

A discrete random variable X and its reduced size-bias $X^!$ have the same distribution if and only if X is a Poisson distribution

Proof.

Solve the differential equation $\frac{\mathrm{d}}{\mathrm{d}t} G_X(t) = \frac{1}{\mu} G_X(t)$, making sure the solution is a valid PGF. $\square$

Poisson as fixed point

Theorem

A discrete random variable X and its reduced size-bias $X^!$ have the same distribution if and only if X is a Poisson distribution

Proof.

Solve the differential equation $\frac{\mathrm{d}}{\mathrm{d}t} G_X(t) = \frac{1}{\mu} G_X(t)$, making sure the solution is a valid PGF. $\square$

This is used in "Stein–Chen approximation":
A random variable is approximately Poisson if X and $X^!$ (or $X^* - 1$) have similar distributions.

An interesting connection

Let X be a positive real-valued random variable.
The **mixed Poisson** distribution $Y \sim \mathrm{MPo}(X)$ means
the conditional distribution. $Y \mid X \sim \mathrm{Po}(X)$.

An interesting connection

Let X be a positive real-valued random variable.
The **mixed Poisson** distribution $Y \sim \mathrm{MPo}(X)$ means
the conditional distribution. $Y \mid X \sim \mathrm{Po}(X)$.

Theorem

Let $Y \sim \mathrm{MPo}(X)$*.*
Then $Y^! \sim \mathrm{MPo}(X^*)$*.*

"The reduced size-bias is the usual size-bias factored through the mixed Poisson operation."

Proof.

Follows from $\mathbb{E}Y = \mathbb{E}X$ and $G_Y(t) = M_X(t-1)$. □

Conclusions

Reasons to prefer the reduced size-bias for discrete random variables:

1. Pleasant examples without "$+1$"s
2. Poisson as fixed point
3. Combines well with factorial moments
4. Combines well with PGF
5. Commutes with thinning

Section 3

Multivariate size-bias

Definition of multivariate size-bias

Definition

Let $\mathbf{X}$ be a random variable whose jth coordinate X_j has expectation $\mu_j = \mathbb{E}X_j$.

The *size-biased version of* X*, biased on coordinate* j, or just the *size-bias on* j of X, is the random vector $\mathbf{X}^{*j}$ such that

$$\mathbb{E}\, X_j\, f(\mathbf{X}) = \mu_j\, \mathbb{E}\, f(\mathbf{X}^{*j})$$

for all "nice" functions $f : \mathbb{R}_+^d \to \mathbb{R}$.

Taking f to be an appropriate indicator function, we get

$$p^{*j}(\mathbf{x}) = \frac{1}{\mu_j}\, x_j\, p(\mathbf{x}).$$

Examples of multivariate size-bias

Examples

Distribution	Size-bias on j
$\Gamma(\nu, \boldsymbol{\lambda})$	$\Gamma(\nu + 1, \boldsymbol{\lambda})$
$\mathrm{Dir}(\boldsymbol{\alpha})$	$\mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{e}_j)$
point mass at $\mathbf{x}$	point mass at $\mathbf{x}$

- $\mathbf{X} \sim \Gamma(\nu, \boldsymbol{\lambda})$ means $X_i = Y/\lambda_i$ where $Y \sim \Gamma(\nu, 1)$.
- Dirichlet distribution is a multivariate generalisation of the Beta with PDF $f(\mathbf{x}) \propto \prod_{i=1}^{d} x_i^{\alpha_i - 1}$, where $\sum_{i=1}^{d} x_i = 1$.

Properties of multivariate size-bias

Theorem

Let $\mathbf{X}$ *be a non-negative random vector with finite* $\mathbb{E}X_j = \mu_j$ *and let* $\mathbf{X}^{*j}$ *be its size-bias on* j.

1. *Expectation:* $\mathbb{E}X_i^{*j} = \mu_i + \dfrac{\mathrm{Cov}(X_i, X_j)}{\mu_j}$
2. *Moments:* $\mathbb{E}(\mathbf{X}^{*j})^{\mathbf{k}} = \dfrac{1}{\mu_j}\,\mathbb{E}\,\mathbf{X}^{\mathbf{k}+\mathbf{e}_j}$
3. *MGF:* $M_{\mathbf{X}^{*j}}(\mathbf{t}) = \dfrac{1}{\mu_j}\dfrac{\mathrm{d}}{\mathrm{d}t_j} M_{\mathbf{X}}(\mathbf{t})$
4. *Size-bias commutes with scaling:* $(a\mathbf{X})^{*j} \equiv a(\mathbf{X}^{*j})$
5. *Size-bias can be done in any order:* $(\mathbf{X}^{*i})^{*j} \equiv (\mathbf{X}^{*j})^{*i}$

Multivariate moments: $\mathbb{E}\,\mathbf{X}^{\mathbf{k}} = \mathbb{E}\prod_{i=1}^{d} X_i^{k_i}$

Coordinate independence

For the multivariate Gamma and point-mass distributions,
the size-bias is the same whichever coordinate you pick.

Coordinate independence

For the multivariate Gamma and point-mass distributions,
the size-bias is the same whichever coordinate you pick.

Theorem

*The size-bias of $\mathbf{X} = (X_i)$ is "coordinate-independent"
if and only if $X_i = a_i Y$ for deterministic $\mathbf{a} = (a_i)$ and univariate Y.*

Section 4

Reduced multivariate size-bias

Definition of reduced multivariate size-bias

Definition

Let $\mathbf{X}$ be a discrete random variable whose jth coordinate X_j has expectation $\mu_j = \mathbb{E}X_j$.

The *reduced size-biased version of X, biased on coordinate j*, or just the *reduced size-bias on j* of X, is the random vector $\mathbf{X}^{!j}$ such that

$$\mathbb{E}\, X_j\, f(\mathbf{X} - \mathbf{e}_j) = \mu_j\, \mathbb{E}\, f(\mathbf{X}^{!j})$$

for all "nice" functions $f : \mathbb{N}^d \to \mathbb{R}$.

Definition of reduced multivariate size-bias

Definition

Let $\mathbf{X}$ be a discrete random variable whose jth coordinate X_j has expectation $\mu_j = \mathbb{E}X_j$.

The *reduced size-biased version of* X*, biased on coordinate* j, or just the *reduced size-bias on* j of X, is the random vector $\mathbf{X}^{!j}$ such that

$$\mathbb{E}\, X_j\, f(\mathbf{X} - \mathbf{e}_j) = \mu_j\, \mathbb{E}\, f(\mathbf{X}^{!j})$$

for all "nice" functions $f : \mathbb{N}^d \to \mathbb{R}$.

Taking f to be an appropriate indicator function, we get

$$p^{!j}(\mathbf{x}) = \frac{1}{\mu_j}\,(x_j + 1)\, p(\mathbf{x} + \mathbf{e}_j).$$

So $X^{!j} \equiv X^{*j} - \mathbf{e}_j$.

Examples of reduced multivariate size-bias

Examples

Distribution	Reduced size-bias on j
$\mathrm{Multi}(n, \mathbf{p})$	$\mathrm{Multi}(n-1, \mathbf{p})$
$\mathrm{NegMulti}(n, \mathbf{q})$	$\mathrm{NegMulti}(n+1, \mathbf{q})$
$\mathrm{Po}(\boldsymbol{\lambda})$	$\mathrm{Po}(\boldsymbol{\lambda})$
$\mathrm{Herm}(\boldsymbol{\mu}, \Sigma)$	$\mathrm{Herm}(\boldsymbol{\mu}, \Sigma) + \mathrm{Bern}\left(\frac{\boldsymbol{\sigma}_j}{\mu_j}\right)$

- $\mathbf{X} \sim \mathrm{Po}(\boldsymbol{\lambda})$ means $X_i \sim \mathrm{Po}(\lambda_i)$ all independent.

Properties of multivariate size-bias

Theorem

Let $\mathbf{X}$ *be a discrete random vector with finite* $\mathbb{E}X_j = \mu_j$ *and let* $\mathbf{X}^{!j}$ *be its reduced size-bias on* k.

1. *Expectation:* $\mathbb{E}X_i^{!j} = \mu_i + \dfrac{\mathrm{Cov}(X_i, X_j)}{\mu_j}$ *for* $i \neq j$
2. *Moments:* $\mathbb{E}(\mathbf{X}^{!j})^{\underline{\mathbf{k}}} = \dfrac{1}{\mu_j}\,\mathbb{E}\,\mathbf{X}^{\underline{\mathbf{k}+\mathbf{e}_j}}$
3. *PGF:* $G_{\mathbf{X}^{!j}}(\mathbf{t}) = \dfrac{1}{\mu_j}\dfrac{\mathrm{d}}{\mathrm{d}t_j}G_{\mathbf{X}}(\mathbf{t})$
4. *Size-bias can be done in any order:* $(\mathbf{X}^{*i})^{!j} \equiv (\mathbf{X}^{!j})^{*i}$

Properties of multivariate size-bias

Theorem

Let $\mathbf{X}$ *be a discrete random vector with finite* $\mathbb{E}X_j = \mu_j$ *and let* $\mathbf{X}^{!j}$ *be its reduced size-bias on* k*.*

1. *Expectation:* $\mathbb{E}X_i^{!j} = \mu_i + \dfrac{\mathrm{Cov}(X_i, X_j)}{\mu_j}$ *for* $i \neq j$
2. *Moments:* $\mathbb{E}(\mathbf{X}^{!j})^{\underline{\mathbf{k}}} = \dfrac{1}{\mu_j}\,\mathbb{E}\,\mathbf{X}^{\underline{\mathbf{k}+\mathbf{e}_j}}$
3. *PGF:* $G_{\mathbf{X}^{!j}}(\mathbf{t}) = \dfrac{1}{\mu_j}\dfrac{\mathrm{d}}{\mathrm{d}t_j} G_{\mathbf{X}}(\mathbf{t})$
4. *Size-bias can be done in any order:* $(\mathbf{X}^{*i})^{!j} \equiv (\mathbf{X}^{!j})^{*i}$

Research question

What the appropriate generalisation of the scaling/thinning result?

Coordinate independence again

For the binomial, negative binomial, and Poisson distributions, the reduced size-bias is the same whichever coordinate you pick.

Coordinate independence again

For the binomial, negative binomial, and Poisson distributions, the reduced size-bias is the same whichever coordinate you pick.

Theorem

The reduced size-bias of $\mathbf{X} = (X_i)$ is "coordinate-independent" if and only if $\mathbf{X}$ is a "mixed multinomial" distribution; that is, if there is some univariate Y and deterministic $\mathbf{a} = (a_i)$ such that

$$\mathbf{X} \mid Y \sim \mathrm{Multi}(Y, \mathbf{a}).$$

Section 5

Size-bias for random measures

Random measures

A **random measure** ξ is a random variable that takes values in the collection $\mathcal{M}(\mathcal{X})$ of measures on a space $\mathcal{X}$.

We assume $\xi(A)$ is finite for bounded sets A.

Random measures

A **random measure** ξ is a random variable that takes values in the collection $\mathcal{M}(\mathcal{X})$ of measures on a space $\mathcal{X}$.

We assume $\xi(A)$ is finite for bounded sets A.

If $\xi(A)$ is a non-negative integer for all bounded sets A,
we call ξ a **point process**,
and think of $\xi(A)$ as "the number of points in A".

Intensity

The equivalent of the expectation of a random variable is the **intensity**.

Here, $\mu(A)$ is the expected measure (or expected number of points in) A.

Definition

Let ξ be a random measure.
The associated *intensity measure* of ξ is the measure μ defined by

$$\mu(A) = \mathbb{E}\,\xi(A).$$

Definition of size-bias for random measures

Definition

Let ξ be a random measure with intensity measure μ, and let B be a bounded set.

The *size-biased version of* ξ, *biased on* B, or just the *size-bias on* B of ξ, is the random measure ξ^{*B} such that

$$\mathbb{E}\,\xi(B)\,f(\xi) = \mu(B)\,\mathbb{E}\,f(\xi^{*B})$$

for all "nice" functions $f : \mathcal{M}(\mathcal{X}) \to \mathbb{R}$.

Compare with the definition of multivariate size-bias

$$\mathbb{E}\,X_j\,f(\mathbf{X}) = \mu_j\,\mathbb{E}\,f(\mathbf{X}^{*j}).$$

Definition of size-bias for random measures

Definition

Let ξ be a random measure with intensity measure μ, and let B be a bounded set.

The *size-biased version of* ξ, *biased on* B, or just the *size-bias on* B of ξ, is the random measure ξ^{*A} such that

$$\mathbb{E}\,\xi(B)\,f(\xi) = \mu(B)\,\mathbb{E}\,f(\xi^{*B})$$

for all "nice" functions $f\colon \mathcal{M}(\mathcal{X}) \to \mathbb{R}$.

Small question

Should I disallow sets B with $\mu(B) = 0$, or should I say $\xi^{*B} \equiv \xi$ if $\mu(B) = 0$?

Examples of size-bias for random measures

Examples

Distribution	**Size-bias on B**
$\Gamma(\nu, \Theta)$	$\Gamma(\nu + 1, \Theta)$
$\mathrm{Dir}(\alpha)$	$\mathrm{Dir}(\alpha_B)$
deterministic ξ	deterministic ξ

- $\xi \sim \Gamma(\nu, \Theta)$ for a deterministic measure Θ means $\xi(A) = Y\,\Theta(A)$ where $Y \sim \Gamma(\nu, 1)$.
- Dirichlet process is the random probability measure version of the Dirichlet distribution, α is a deterministic measure, and

$$\alpha_B(A) = \alpha(A) + \frac{\alpha(A \cap B)}{\alpha(B)}.$$

Properties of size-bias for random measures

Theorem

Let ξ be a random measure with intensity μ and ξ^{*B} be its size-bias on B.

1. *Intensity:* $\mu^{*B}(A) = \mu(A) + \dfrac{\mathrm{V}_\xi(A,B)}{\mu(B)}$
2. *Moment measures:* $m_k^{*B}(E) = \dfrac{1}{\mu(B)}\, m_{k+1}(E \times B)$ *for* $E \subset \mathcal{X}^k$
3. *Laplace functional:* $L_{\xi^{*B}}(u) = \dfrac{1}{\mu(B)}\, \dfrac{\delta}{\delta \mathbf{1}_B}\, L_\xi(u)$
4. *Size-bias commutes with scaling:* $(a\xi)^{*B} \equiv a(\xi^{*B})$

- $\mathrm{V}_\xi(A,B) = \mathrm{Cov}(\xi(A), \xi(B))$
- m_k are the moment measures on $\mathcal{X}^k$, the equivalent of the moments.
- The Laplace functional is the equivalent of the MGF, and $\frac{\delta}{\delta \mathbf{1}_B}$ is a functional derivative.

Fixed point: deterministic measures

Conjecture

The size-bias ξ^{*B} has the same distribution as ξ for all B if and only if ξ is a deterministic measure.

Fixed point: deterministic measures

Conjecture

The size-bias ξ^{*B} has the same distribution as ξ for all B
if and only if ξ is a deterministic measure.

Research questions

- What are the correct restrictions on B?
  (Just "bounded and measurable"?)
- Do I need to put "ξ-almost everywhere" statements in some places to make this true?
- What's the best way to prove this?

Section 6

Reduced size-bias for point processes

Definition of reduced size-bias for point process

Again, a "reduced" definition works better for discrete random measures – i.e. point processes.

Definition of reduced size-bias for point process

Again, a "reduced" definition works better for
discrete random measures – i.e. point processes.

Definition

Let ξ be a point process with intensity measure μ,
and let B be a bounded set.

The *reduced size-biased version of* ξ*, biased on* B, or just
the *reduced size-bias on* B of ξ is the point process $\xi^{!B}$ such that

$$\mathbb{E} \int_B f(\xi - \delta_x)\, \xi(\mathrm{d}x) = \mu(B)\, \mathbb{E}\, f(\xi^{!B})$$

for all "nice" functions $f \colon \mathcal{M}^{\#}(\mathcal{X}) \to \mathbb{R}$.

Explaining the definition

Definition

$$\mathbb{E} \int_B f(\xi - \delta_x)\, \xi(\mathrm{d}x) = \mu(B)\, \mathbb{E}\, f(\xi^{!B})$$

Informally,

$$\mathbb{E} \int_B f(\xi - \delta_x)\, \xi(\mathrm{d}x) = \mathbb{E} \sum_{X \in \xi \cap B} f(\xi - \delta_X),$$

where the sum is over the points X of ξ within B.

Explaining the definition

Definition

$$\mathbb{E} \int_B f(\xi - \delta_x)\, \xi(\mathrm{d}x) = \mu(B)\, \mathbb{E}\, f(\xi^{!B})$$

Informally,

$$\mathbb{E} \int_B f(\xi - \delta_x)\, \xi(\mathrm{d}x) = \mathbb{E} \sum_{X \in \xi \cap B} f(\xi - \delta_X),$$

where the sum is over the points X of ξ within B.

Compare to the left-hand side of the standard size-bias definition

$$\mathbb{E}\, \xi(B)\, f(\xi) = \mathbb{E} \sum_{X \in \xi \cap B} f(\xi).$$

Examples of reduced size-bias for point processes

Examples

Distribution	Reduced size-bias on j
$\mathrm{BP}(n, \Theta)$	$\mathrm{BP}(n-1, \Theta)$
$\mathrm{NBP}(n, \Theta)$	$\mathrm{NBP}(n+1, \Theta)$
$\mathrm{PP}(\Lambda)$	$\mathrm{PP}(\Lambda)$
$\mathrm{GPP}(\mu, \Sigma)$	$\mathrm{GPP}(\mu, \Sigma) + \mathrm{BP}\left(1, \frac{\Sigma(\mathcal{B}\times\cdot)}{\mu(B)}\right)$

- GPP is the "Gauss–Poisson process", a clustered Poisson process that has similar properties to the Hermite distribution.

Properties of reduced size-bias for point processes

Theorem

Let ξ be a point process with intensity μ and $\xi^{!B}$ be its reduced size-bias.

1. *Intensity:* $\mu^{!B}(A) = \mu(A) + \dfrac{\mathrm{D}_\xi(A, B)}{\mu(B)}$
2. *Factorial moment measures:* $m^{!B}_{\underline{k}}(E) = \dfrac{1}{\mu(B)}\, m_{\underline{k+1}}(E \times B)$
3. *PGFl:* $G_{\xi^{!B}}(u) = \dfrac{1}{\mu(B)}\, \dfrac{\delta}{\delta \mathbf{1}_B}\, G_\xi(u)$
4. *Size-bias commutes with thinning:* $(a \circ \xi)^{!B} \equiv a \circ (\xi^{!B})$

- $\mathrm{D}_\xi(A, B) = \mathrm{Cov}(\xi(A), \xi(B)) - \mathbb{E}\,\xi(A \cap B)$ is like the dispersion
- $m_{\underline{k}}$ is to m_k as $\mathbb{E}\, X^{\underline{k}}$ is to $\mathbb{E}\, X^k$.
- G_ξ is the probability generating functional
- Thinning means delete each point with probability $1 - a$

Fixed point for point processes

Conjecture

The reduced size-bias ξ^{*B} has the same distribution as ξ for all B if and only if ξ is a Poisson process.

Research questions

Again, what is the most accurate and general way to state and prove this?

Set independence for point processes

Conjecture

The reduced size-bias ξ^{*B} is the same for all sets B
if and only if ξ is a mixed Poisson process;
that is, if $\xi \mid Y \sim \mathrm{PP}(Y\Lambda)$ for univariate random Y and deterministic measure Λ.

Research questions

Is this true? what is the most accurate and general way to state this?

Connection with Cox processes

We have a result similar to the one
about mixed Poisson distributions earlier.

If ξ is a random measure, the **Cox process** $\eta \sim \mathrm{Cox}(\xi)$ is, conditional on ξ, a Poisson process with intensity ξ.

Connection with Cox processes

We have a result similar to the one
about mixed Poisson distributions earlier.

If ξ is a random measure, the **Cox process** $\eta \sim \mathrm{Cox}(\xi)$ is, conditional on ξ, a Poisson process with intensity ξ.

Theorem

Let $\eta \sim \mathrm{Cox}(\xi)$.
Then $\eta^{!B} \sim \mathrm{Cox}(\xi^{*B})$.

Section 7

Connections with Palm theory

What is the Palm distribution?

Let ξ be a point process.
Informally, the **Palm process** $\xi^{\mathsf{P}x}$ at $x \in \mathcal{X}$ is what you get by conditioning on there being a point at location x.

What is the Palm distribution?

Let ξ be a point process.
Informally, the **Palm process** $\xi^{\mathsf{P}x}$ at $x \in \mathcal{X}$ is what you get by conditioning on there being a point at location x.

The formal definition is...

Definition

Let ξ be a random measure. Then the collection $(\xi^{\mathsf{P}x} : x \in \mathcal{X})$ of *Palm measures at locations x* (collectively the *"Palm kernel"*) is defined by

$$\mathbb{E} \int_{\mathcal{X}} g(x, \xi)\, \xi(\mathrm{d}x) = \int_{\mathcal{X}} \mathbb{E}\, g\left(x, \xi^{\mathsf{P}x}\right) \mu(\mathrm{d}x).$$

for all "nice" functions $g \colon \mathcal{X} \times \mathcal{M}(\mathcal{X}) \to \mathbb{R}$.

...although it's not clear why this matches the informal explanation.

Conditioning and size-bias

If the point process ξ and the set B are such that
we know that $\xi(B) \in \{0, 1\}\dots$

$\dots$ then the size-bias ξ^{*B} is equivalent to conditioning that $\xi(B) = 1$.

"If there is at are either zero or one points in B
then size-bias is like conditioning there is one."

Conditioning and size-bias

If the point process ξ and the set B are such that we know that $\xi(B) \in \{0, 1\}$...

...then the size-bias ξ^{*B} is equivalent to conditioning that $\xi(B) = 1$.

"If there is at are either zero or one points in B then size-bias is like conditioning there is one."

This feels a lot like the Palm process with $B = \{x\}$...

Conditioning and size-bias

If the point process ξ and the set B are such that
we know that $\xi(B) \in \{0, 1\}\ldots$

$\ldots$ then the size-bias ξ^{*B} is equivalent to conditioning that $\xi(B) = 1$.

"If there is at are either zero or one points in B
then size-bias is like conditioning there is one."

This feels a lot like the Palm process with $B = \{x\}\ldots$

$\ldots$ but this is not allowed, since $\mu(B)$ must be non-zero.

From size-bias to Palm process

If you take a sequence of non-null sets $B_n \to \{x\}$,
then, at least informally, we have $\xi^{*B_n} \to \xi^{\mathsf{P}x}$,
in that the formulas in the definitions converge.

From size-bias to Palm process

If you take a sequence of non-null sets $B_n \to \{x\}$,
then, at least informally, we have $\xi^{*B_n} \to \xi^{\mathrm{P}x}$,
in that the formulas in the definitions converge.

Research question

Can this be made rigorous?

From Palm distribution to size-bias

An appropriate average of the Palm measures gives the size-bias, in the following sense:

Theorem

Let ξ be a random measure, $(\xi^{\mathsf{P}x})$ its Palm kernel, and ξ^{*B} its size-bias on B. Then

$$\mathbb{E}\, f\left(\xi^{*B}\right) = \frac{1}{\mu(B)} \int_B \mathbb{E}\, f\left(\xi^{\mathsf{P}x}\right) \mu(\mathrm{d}x).$$

Reduced Palm process and size-bias

There is also a **reduced Palm process** $\xi^{\mathsf{P}!x}$: condition on there being a point at location x, then remove that point.

$\xi^{\mathsf{P}!x}$ corresponds to $\xi^{!B}$ in the same way that $\xi^{\mathsf{P}x}$ corresponds to ξ^{*B}.

I would like to know...

Research question

To what extent are the results I have discussed for size-bias already known in the context of Palm theory?

Palm-first or size-bias-first?

Advantages of the "size-bias first" approach
over the "Palm-first approach":

1. More obvious how the informal explanation leads to the formal definition.
2. Definition tells us about the specific ξ^{*B} we are interested in rather than the kernel of all the $\xi^{\mathsf{P}x}$'s simultaneously.
3. As with conditional probability, we should define the $\mu(B) > 0$ version first.

Section 8

A final research question

Section 9

In conclusion. . .

Conclusions

1. For discrete random variables, the reduced size-bias $X^! = X^* - 1$ has surprisingly pleasant and underexplored properties.
2. There is a multivariate generalisation that seems to have got no attention.
3. We can also define size-bias and reduced size-bias for random measures and point processes.
   They are a bit like Palm measures/processes.